

UNIVERSIDAD MIGUEL HERNÁNDEZ DE ELCHE

ESCUELA POLITÉCNICA SUPERIOR DE ELCHE

GRADO EN INGENIERÍA ELECTRÓNICA Y
AUTOMÁTICA INDUSTRIAL



UNIVERSITAS
Miguel Hernández

"EVALUACIÓN DE DESCRIPTORES
VISUALES EN ENTORNOS CON CAMBIOS
DE ILUMINACIÓN PARA LA
NAVEGACIÓN DE ROBOTS MÓVILES "

TRABAJO FIN DE GRADO

Junio -2024

AUTOR: Marina Pastor Amat

DIRECTOR/ES: Mónica Ballesta Galdeano

Miriam Máximo Gutiérrez

Índice de contenido

1 INTRODUCCIÓN.....	6
2 ROBÓTICA MÓVIL	6
2.1 Navegación autónoma	9
2.2 Tipos de sensores.....	10
3 MARCAS VISUALES.....	11
3.1 Marcas naturales y artificiales.....	12
3.2 Detección y descripción	14
3.2.1. MÉTODOS DE CARACTERÍSTICAS LOCALES.....	17
A) REDES NEURONALES.....	25
3.2.2. MÉTODOS DE CARACTERÍSTICAS GLOBALES.....	31
4 ESTUDIO DE DESCRIPTORES VISUALES ANTE CAMBIOS DE ILUMINACIÓN.....	35
4.1 – OBJETIVOS.....	35
4.2 – MÉTRICAS PARA EVALUAR LOS EXPERIMENTOS REALIZADOS.....	36
4.3 – EXPERIMENTOS Y RESULTADOS	42
4.3.1. Métodos de características locales.....	47
4.3.2. Métodos de características globales.....	74
5- CONCLUSIONES	79
6. REFERENCIAS	82

Índice de figuras y tablas.

Figura 1. Robot RB-WATCHER.....	7
Figura 2. Robot de entrega última milla.....	8
Figura 3. Josie Pepper, robótica asistente del aeropuerto de Múnich.....	9
Figura 4. Interrelación entre los conceptos que engloba la navegación móvil.....	10
Figura 5. Las figuras muestran los distintos tipos de marcas. La Figura (a) son AprilTags en objetos y lugares. Y la Figura (b) son marcas visuales naturales en exterior, imagen tomada en el Campus de la UMH.....	13
Figura 6. Imagen con puntos detectados y descripción del punto X1.....	15
Figura 7. Ejemplo de clustering en 3 puntos seguidos.....	16
Figura 8. Descriptores SIFT.....	19
Figura 9. Puntos SURF.....	20
Figura 10. Descriptor SURF.....	22
Figura 11. Resultado de coincidencias utilizando ORB.....	22
Figura 12. Resultado tras utilizar BRISK.....	24
Figura 13. Coincidencias utilizando AKAZE.....	25
Figura 14. Imagen que “observa” la red neuronal en Autopilot de Tesla.....	27
Figura 15. Estructura de capas de la red SuperPoint [21].	29
Figura 16. Estructura de capas de la red D2-NET [22].	30
Figura 17. Costo computacional del método Firma de Fourier.....	33
Figura 18. En la figura 18.a. tenemos la imagen HOG. En la figura 18.b. tenemos las celdas y el histograma HOG.....	34
Figura 19. La secuencia a) es la número 1 de la base de datos BASLER-exterior. Y la secuencia b) es la número 2 de la base de datos ARVC de interior.....	36
Figura 20. Tabla con el número de secuencias e imágenes interiores.....	42
Figura 21. Tabla con el número de secuencias e imágenes exteriores.....	42
Figura 22.. Imágenes 1,8,14 y 27 de la secuencia 1 de interior - BASLER.....	43
Figura 23. Imágenes 1,4,9 y 14 de la secuencia 1 de exterior - BASLER.....	43
Figura 24. Imágenes 1,6,12 y 20 de la secuencia 2 de exterior - BASLER.....	43
Figura 25.. Imágenes 1,4,9 y 13 de la secuencia 1 de interior- ARVC.....	44
Figura 26.. Imágenes 1,3,6 y 9 de la secuencia 2 de interior - ARVC.....	44
Figura 27. Imágenes 1,4,7 y 9 de la secuencia 3 de interior - ARVC.....	44
Figura 28. Imágenes 1,5,10 y 15 de la secuencia 1 de exterior - ARVC.....	45
Figura 29. Imágenes 1,4,11 y 15 de la secuencia 2 de exterior - ARVC.....	45
Figura 30. Imágenes 1,3,7 y 14 de la secuencia 3 de exterior - ARVC.....	45
Figura 31. Imágenes 1,7,11 y 16 de la secuencia 4 de exterior - ARVC.....	46
Figura 32. Imágenes 1,6,11 y 17 de la secuencia 6 de exterior- ARVC.....	46
Figura 33. Imágenes 1,4,8 y 15 de la secuencia 7 de exterior - ARVC.....	46
Figura 34. Puntos característicos y número de clusters obtenidos para una secuencia de interior (BASLER).	47
Figura 35. Ratio de repetibilidad - BASLER- database - indoor -sec1.....	48
Figura 36. Ratio de repetibilidad - BASLER- database - outdoor -sec1.....	48
Figura 37. Ratio de repetibilidad - BASLER- database - outdoor -sec2.....	49
Figura 38. Ratio de repetibilidad - ARVC- database - indoor -sec1.....	49
Figura 39. Ratio de repetibilidad - ARVC- database - indoor -sec2.....	50
Figura 40. Ratio de repetibilidad - ARVC- database - indoor -sec3.....	50
Figura 41. Ratio de repetibilidad - ARVC- database - outdoor -sec1.....	50
Figura 42. Ratio de repetibilidad - ARVC- database - outdoor -sec2.....	51

Figura 43. Ratio de repetibilidad - ARVC- database - outdoor -sec3	51
Figura 44. Ratio de repetibilidad - ARVC- database - outdoor -sec4	51
Figura 45. Ratio de repetibilidad - ARVC- database - outdoor -sec6	52
Figura 46. Ratio de repetibilidad - ARVC- database - outdoor -sec7	52
Figura 47. Tabla con el número de clusters y porcentaje de acierto. Para secuencia interior de BASLER	54
Figura 48. Recall vs Precision, BASLER-database/indoor/sec1	55
Figura 49. Recall vs Precision, BASLER-database/outdoor/sec1	55
Figura 50. Recall vs Precision, BASLER-database/outdoor/sec2	56
Figura 51. Recall vs Precision, ARVC-database/indoor/sec1	56
Figura 52. Recall vs Precision, ARVC-database/indoor/sec2	57
Figura 53. Recall vs Precision, ARVC-database/indoor/sec3	57
Figura 54. Recall vs Precision, ARVC-database/outdoor/sec1	57
Figura 55. Recall vs Precision, ARVC-database/outdoor/sec2	58
Figura 56. Recall vs Precision, ARVC-database/outdoor/sec3	58
Figura 57. Recall vs Precision, ARVC-database/outdoor/sec4	58
Figura 58. Recall vs Precision, ARVC-database/outdoor/sec6	59
Figura 59. Recall vs Precision, ARVC-database/outdoor/sec7	59
Figura 60. Tabla (a) y gráfica (b) parámetros separabilidad. BASLER-database/indoor/sec1	62
Figura 61. Tabla (a) y gráfica (b) parámetros separabilidad. BASLER-database/outdoor/sec1	63
Figura 62. Tabla (a) y gráfica (b) parámetros separabilidad. BASLER-database/outdoor/sec2	64
Figura 63. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/indoor/sec1	65
Figura 64. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/indoor/sec2	66
Figura 65. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/indoor/sec3	67
Figura 66. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/outdoor/sec1	68
Figura 67. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/outdoor/sec2	69
Figura 68. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/outdoor/sec3	70
Figura 69. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/outdoor/sec4	71
Figura 70. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/outdoor/sec6	72
Figura 71. Tabla (a) y gráfica (b) parámetros separabilidad. ARVC-database/outdoor/sec7	73
Figura 72. Curva Recall vs Precision BASLER-database/indoor/sec1, 71.a) HOG, 71.b) FOURIER	74
Figura 73. Curva Recall vs Precision BASLER-database/outdoor/sec1-2, 72.a) HOG, 72.b) FOURIER	75
Figura 74. Curva Recall vs Precision ARVC-database/indoor/sec1-3, 73.a) HOG, 73.b) FOURIER	75
Figura 75. Curva Recall vs Precision ARVC-database/outdoor/sec1-7, 74.a) HOG, 74.b) FOURIER	75

Figura 76. Tabla de parámetros del criterio de separabilidad. BASLER- indoor- sec1 77
Figura 77. Tabla de parámetros del criterio de separabilidad. BASLER- outdoor- sec1-2
..... 77
Figura 78. Tabla de parámetros del criterio de separabilidad. ARVC- indoor- sec1-3. 78
Figura 79. Tabla de parámetros del criterio de separabilidad. ARVC- outdoor- sec 1-7
..... 78



1 INTRODUCCIÓN

En este trabajo se propone estudiar el comportamiento de una serie de descriptores locales y globales, que incluyen redes neuronales. Estos descriptores están basados en algoritmos cuyo objetivo es codificar la apariencia de un entorno con la información más característica de las imágenes captadas en dicho entorno. En concreto, se pretende analizar la invariabilidad de estos métodos ante cambios de iluminación en la escena. Para ello, se utilizarán varias secuencias de imágenes y diferentes métricas de evaluación, con el fin de determinar qué descriptor visual puede resultar más efectivo o de mayor interés para la navegación de robots autónomos.

2 ROBÓTICA MÓVIL

Los robots actuales son obras de ingeniería concebidas para producir bienes y servicios o explotar recursos naturales [1]. Hasta ahora, el desarrollo de las máquinas ha estado influido por el desarrollo tecnológico, por lo que se parte de máquinas que tienen como objetivo único ampliar la potencia muscular del hombre sustituyéndolo en su trabajo físico a máquinas que son capaces de procesar información complementando e incluso sustituyendo al hombre en actividades intelectuales. Por otra parte, desde la antigüedad, el hombre se ha interesado por las máquinas que imitan a la figura y los movimientos de los seres animados. Desde el mundo griego hasta la actualidad, pasando por los artesanos franceses y suizos del siglo XVIII.

La robótica móvil, es un campo de la robótica que se centra en desarrollar robots que sean autónomos y capaces de operar en entornos variables. Esta disciplina surge de la necesidad de ampliar la autonomía de los robots limitando todo lo posible la intervención humana. Esta autonomía se basa en un sistema de navegación automática, incluyendo tareas de planificación, percepción y control. Como hemos comentado, la presencia de robots móviles en la industria y en ambientes domésticos ha incrementado en los últimos años. Esto ha sido impulsado por varios factores. en cuanto a la industria, la automatización de procesos cada vez está más demandada, usando por ejemplo robots de almacén

o vehículos guiados. También, la reducción de costes y el aumento de la eficiencia ha promovido el uso de los robots móviles.

Algunos ejemplos de aplicaciones de la robótica móvil en el ámbito industrial son:
[2]

- **Inspección y mantenimiento:** robots móviles equipados con sensores de alta precisión como el de la Figura 1, RB-WATCHER, realizan inspecciones precisas en entornos industriales, identificando anomalías en equipos, infraestructuras o productos. También pueden monitorear continuamente condiciones ambientales como temperatura, humedad o niveles de contaminación.



Figura 1. Robot RB-WATCHER.

- **Transporte de materiales:** se trata de uno de los usos más extendidos es transportar materiales en entornos logísticos de fabricación o almacenes, tanto en interiores como exteriores. Estos robots pueden transportar cargas pesadas a lo largo de rutas predefinidas o adaptarse dinámicamente a cambios en la disposición de la planta.
- **Manipulación móvil:** los robots de manipulación móvil ofrecen como principal ventaja la posibilidad de que el brazo robótico industrial pueda ejecutar tareas de manipulación de piezas y materiales en diferentes ubicaciones dentro de una misma planta de trabajo. Esta movilidad facilita la flexibilidad y evita los costes de

las instalaciones fijas que suponen los brazos robóticos industriales.

- **Entrega de última milla:** estos robots autónomos agilizan la distribución de paquetes en áreas urbanas, navegando por aceras y calles con precisión. Integran sistemas de seguimiento y planificación de rutas para garantizar entregas rápidas. Podemos apreciar un ejemplo en la Figura 2.



Figura 2. Robot de entrega última milla.

- **Agricultura.** En el sector agrícola, los robots autónomos automatizan tareas como la siembra, cosecha, riego y control de plagas. La monitorización precisa del estado del cultivo contribuye a una producción más sostenible y rentable. Están equipados con herramientas especializadas como brazos robóticos y sistemas de visión.
- **Construcción e infraestructuras.** Los robots de construcción realizan tareas como el vertido de hormigón, corte y soldadura de materiales o inspección de infraestructuras civiles. Operan en entornos peligrosos o de difícil acceso, reduciendo el riesgo para los trabajadores humanos. Agilizan el proceso de construcción y aceleran el tiempo de finalización del proyecto.

Por otro lado, en cuanto al ámbito doméstico, los robots de limpieza siguen disparando las ventas, así como los asistentes personales/robots de compañía o los robots de cocina. También están siendo utilizados para seguridad y vigilancia y para entretenimiento y educación. Estos asistentes personales, están siendo utilizados en lugares públicos, como Josie Pepper, la robótica asistente personal del aeropuerto de Múnich [3]. Este robot humanoide se mueve con

rodillos, tiene voz, ojos y se encarga de resolver las dudas de los pasajeros del aeropuerto. Está equipado con una inteligencia artificial muy desarrollada, la cual permite a Josie intentar resolver cada pregunta individualmente y no con textos genéricos, por lo que hace que se asemeje mucho a un cerebro humano.



Figura 3. Josie Pepper, robótica asistente del aeropuerto de Múnich.

2.1 Navegación autónoma

La autonomía de un robot, es el grado más complejo que puede alcanzar un robot [4]. Para llevar a cabo esta tarea, es muy importante la capacidad de percibir el entorno, para ello los robots son dotados de sensores que les permiten extraer la información necesaria para construir modelos del entorno.

Un robot móvil es un robot con capacidad de locomoción que se basa en saber dónde está, a dónde quiere ir y en cómo llegar a su destino. A raíz de estas cuestiones surgen tres problemas fundamentales a los que se enfrenta la robótica móvil: “Localización”, “*Mapping*” y “*Path Planning*”

- Localización. La localización hace referencia a la estimación de la pose del robot, y esto se consigue a partir de los datos de los sensores y de un mapa
- *Mapping*. El *mapping* se refiere a la construcción de un mapa, interpretando los datos de los sensores que construyen un modelo del entorno. Para ello se considera que se conoce la posición del robot en todo momento

- *Path Planning*. El *Path Planning* o planificación de trayectorias, estudia cuál es el mejor camino para ir de un lugar a otro.

Estos conceptos se relacionan entre sí como podemos apreciar en la Figura 4, ya que como hemos dicho para construir un mapa del entorno, es necesario conocer la posición del robot, por lo que los problemas de *mapping* y *localización* están relacionados.

Para resolver esta problemática de forma simultánea surge SLAM, *Simultaneous Localization and Mapping*, o CMBL, *Concurrent Map Building and Localization*.

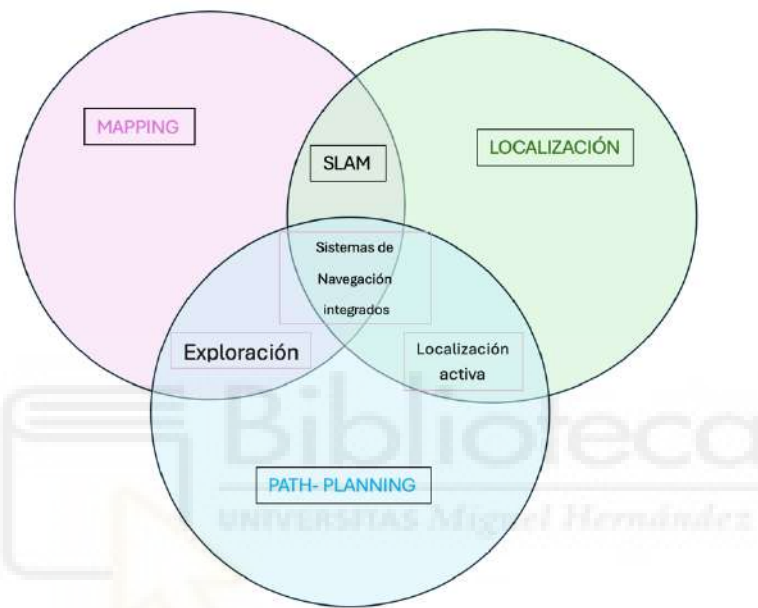


Figura 4. Interrelación entre los conceptos que engloba la navegación móvil

2.2 Tipos de sensores.

Además de estos conceptos, y de otros como el control del movimiento o la detección y evitación de obstáculos, como hemos comentado, la percepción del entorno es muy importante para conseguir la navegación autónoma, y para ello se utilizan los sensores.

Los sensores pueden ser de tipología muy variada [5], como, por ejemplo

- **Sensores de luz.** Se utilizan para identificar la luz y generar una diferencia de tensión. Hay dos tipos que se usan principalmente: fotorresistencias y células fotovoltaicas.
- **Sensores sonoros.** Se utiliza para reconocer un sonido y convertirlo en una señal eléctrica. Se usa en robots simples capaces de navegar con la ayuda del sonido.
- **Sensores de temperatura.** Se utilizan mucho en robots que trabajan en condiciones extremas, de manera que los sensores le ayudan a adaptarse y controlar su funcionamiento
- **Sensores de contacto.** Los sensores de contacto dependen de la interacción física para funcionar. Son muy utilizados para que los robots eviten obstáculos, de manera que, si el sensor del robot toca un objeto, ordena al robot realizar una determinada acción como girar o detenerse. Un ejemplo de sensor de contacto muy utilizado en el mercado es el sensor de fin de carrera.
- **Sensores de Proximidad.** Estos sensores pueden detectar la presencia de un objeto dentro de distancias establecidas, sin necesidad de contacto. Algunos de los más comunes son los transmisores de infrarrojos y los ultrasónicos. También se puede utilizar como sensor de proximidad, las fotorresistencias mencionadas en los sensores de luz.
- **Sensores de navegación.** Los sensores de navegación o posicionamiento se usan en robots para detectar su posición en el mundo. Los más comunes son GPS, Localización y Brújula.
- **Sensores ópticos.** El sistema de visión de los robots consiste en una o más cámaras conectadas a un computador, éste posee un software de procesamiento que ayuda al robot a interpretar lo que la cámara ve.

3 MARCAS VISUALES

La construcción de mapas comienza a partir de la información visual del entorno que percibe el robot gracias al sistema de visión que posee. Este sistema está formado por sensores, como pueden ser cámaras, que le permiten extraer puntos característicos de la escena.

La posición de dichos puntos se define mediante un sistema de coordenadas global, y cada punto está caracterizado mediante información visual de su entorno local. La posición del punto y su descriptor visual asociado componen una marca visual (visual *landmark*) [6].

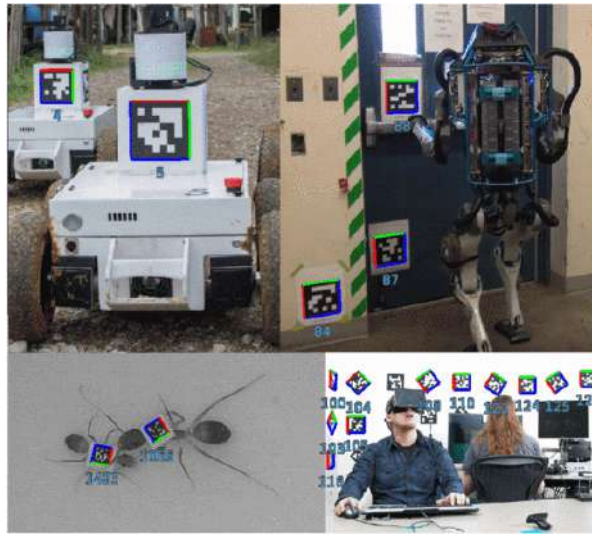
Por lo que el mapa se construye mediante un conjunto de marcas visuales que son integradas y actualizadas a lo largo del proceso de SLAM.

3.1 Marcas naturales y artificiales

Podemos encontrar dos tipos de *landmarks*, Artificiales y Naturales. Las marcas artificiales consisten en características añadidas al entorno “artificialmente”, es decir, por el hombre, con el objetivo de facilitar la navegación del robot. Por ejemplo, podría añadirse códigos de barras al entorno, para que el robot los pueda identificar, y de esta manera la localización del robot móvil se simplifica extremadamente. Ejemplos de marcas visuales artificiales usadas hoy en día son los códigos QR y marcadores ART, ArUco, STag o AprilTag [7]. Los AprilTag, consisten en características visuales artificiales diseñadas para la detección automática, a menudo llevan una carga útil para que sean fácilmente distinguibles entre sí, como podemos ver en la figura 5.a. Sus usos van desde la verdad en tierra hasta la detección y el seguimiento de objetos. Su singularidad y alta tasa de detección son ideales para probar sistemas SLAM. Por otra parte STag [8], está diseñado para ser robusto contra los factores de fluctuación, por lo que mantiene la estabilidad de la pose.

Por otra parte, cuando hablamos de marcas naturales, nos referimos a marcas que ya posee el entorno sin la intervención humana. Por ejemplo, en entornos cerrados o interiores pueden ser esquinas, puertas o paredes, ya que se pueden identificar fácilmente en el mapa

del entorno. En exteriores, podrían ser los troncos de los árboles. En la Figura 5.b podemos ver un ejemplo de estas marcas.



(a)



(b)

Figura 5. Las figuras muestran los distintos tipos de marcas. La Figura (a) son AprilTags en objetos y lugares. Y la Figura (b) son marcas visuales naturales en exterior, imagen tomada en el Campus de la UMH.

Como hemos comentado, definimos una marca visual como un punto en el espacio que es fácilmente detectable utilizando imágenes del entorno, y a continuación nombraremos una serie de características que son las deseables en una marca visual.

- Debe ser fácilmente detectable. Por lo que el robot debe ser capaz de detectarla desde diferentes posiciones del entorno
- Debe ser fácilmente reconocible. Esto implica que la apariencia visual de la marca debe ser descrita de manera que el robot pueda distinguir entre diferentes puntos en el espacio. Esta descripción debe ser invariante ante cambios en el punto de vista y la distancia, ya que el robot puede observar la marca visual desde distintas posiciones del entorno.
- Debe situarse sobre puntos característicos del entorno por el que se mueve el robot. Si hay zonas donde no hay carencia de detalle o textura no hay marcas visuales.
- Se debe poder ubicar con precisión en el espacio. Esto surge de la necesidad de obtener medidas de distancia precisas utilizando sistemas de visión.

3.2 Detección y descripción

El proceso que realiza el robot explorando el entorno consiste en tomar imágenes, encontrar marcas visuales en ellas y construir el mapa del entorno. Durante el proceso de encontrar las marcas visuales, surgen dos procesos fundamentales:

- Detección: este proceso implica procesar las imágenes tomadas del entorno, de manera que se obtienen una serie de puntos de interés. El objetivo de este proceso es obtener una serie de puntos en el entorno que sean estables y que puedan ser detectados desde diferentes ángulos y distancias ante cambios de posición y orientación de la cámara
- Descripción: la descripción de una marca visual consiste en construir un vector de características calculado en base a la apariencia visual del punto en el espacio. La descripción buscará ser invariante ante cambios en la posición y orientación con la que la marca es vista

Podemos apreciar un ejemplo de este proceso en la Figura 6, donde podemos ver en la imagen los puntos de interés obtenidos mediante un método de detección, en concreto nos centramos en $X1$, el cual se describe con un método de descripción, quedando de la siguiente manera: $X1 = \{d1, d2, d3, \dots, dn\}$ donde n , es el número de descriptores.

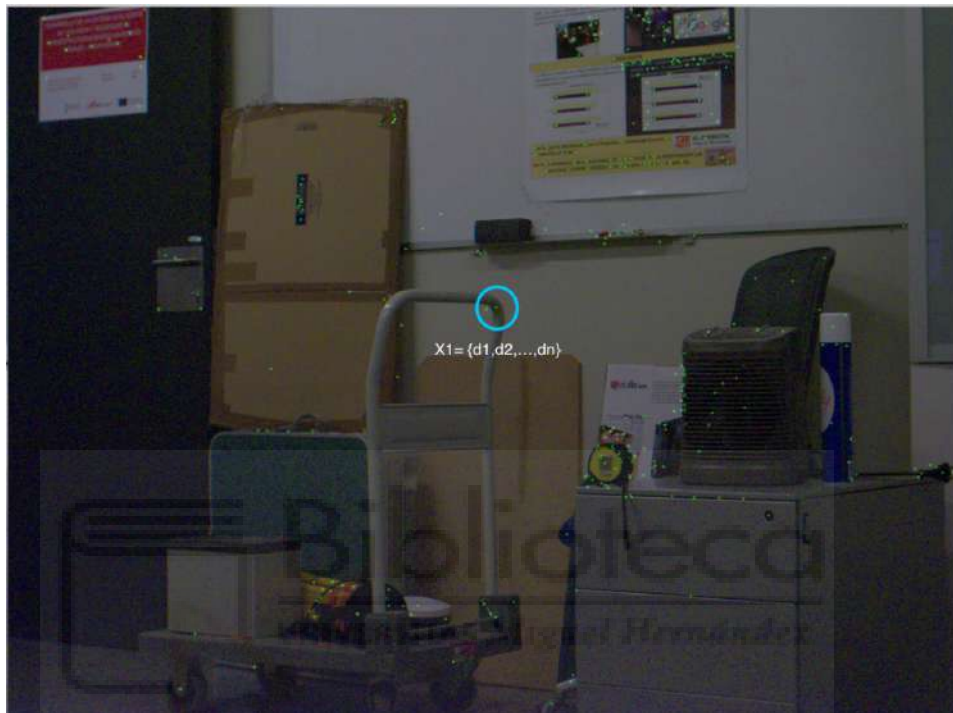


Figura 6. Imagen con puntos detectados y descripción del punto $X1$

Un proceso de estudio es evaluar la estabilidad e invariabilidad de los puntos de interés y la capacidad que tienen los descriptores de caracterizar dichos puntos a lo largo de secuencias con diferentes variaciones en la imagen, como cambios de puntos de vista o de ángulo e iluminación. En nuestro caso lo estudiaremos para variaciones de iluminación.

Esta situación recrea las diferentes posiciones desde las que el robot observa un punto cuando realiza tareas de SLAM. En nuestro caso, en vez de tener pares de puntos correspondientes entre las imágenes como en otros estudios, decimos que una marca visual está representada por un conjunto (*cluster*) de descriptores extraídos desde diferentes

posiciones. El conjunto se compone de un punto que ha sido observado desde diferentes posiciones en una secuencia de imágenes y tiene asociado el descriptor en cada imagen. Cada *cluster* representa el punto detectado en las imágenes donde ha sido encontrado. Por ejemplo, en la Figura 7, podemos observar tres puntos (X_1, X_2 y X_3) que han sido seguidos en dos imágenes con distinta iluminación, y se muestran ejemplos de *clustering* de los puntos seguidos.



Figura 7. Ejemplo de clustering en 3 puntos seguidos

En siguiente lugar se procede a realizar el seguimiento (tracking) de los puntos detectados. Como hemos comentado anteriormente, trabajaremos con *clusters*, que son conjuntos que se componen de puntos que han sido observados desde diferentes posiciones en una secuencia de imágenes, como podemos apreciar en la Figura 7, donde se siguen los puntos x_1 , x_2 y x_3 , en las dos imágenes de la secuencia. Si las imágenes fuesen capturadas por el robot en movimiento, habría desplazamiento de los objetos y habría que establecer un método geométrico de seguimiento de puntos. En nuestro caso se trata de imágenes estáticas, por lo que no es necesario realizar este seguimiento. Lo que haremos será buscar los puntos detectados en una imagen de la secuencia, en la siguiente imagen de la secuencia, pero acotando una región.

En el presente trabajo, únicamente contemplaremos secuencias con cambios de iluminación, las cuales son estáticas en cuanto a la posición del punto de observación, por lo que como hemos comentado no será necesario aplicar estos métodos. Este proceso lo evaluaremos mediante

la métrica de evaluación “Ratio de repetibilidad” que describiremos más adelante en el apartado 4.2

3.2.1. MÉTODOS DE CARACTERÍSTICAS LOCALES

A continuación, describiremos una serie de métodos de extracción de marcas visuales los cuales son detectores y descriptores. Existen otros métodos muy comunes como *Harris Corner Detector*, el cual es sólo detector y se utilizó para extraer marcas visuales en SLAM monocular [16]. Otro método de detección es *SUSAN (Smallest Univalve Segment Assimilating Nucleus)*, y era utilizado tradicionalmente en aplicaciones de reconocimiento de objetos [12]. Por otra parte, también se han aplicado métodos de descripción anteriormente, como *Ventana de niveles de grises*, utilizado también en un contexto de SLAM monocular [16] o *Momentos de Zernike*, presentado con la formulación de los polinomios de Zernike [18], siendo uno de los más populares en términos de redundancia y capacidad de reconstrucción.

En este trabajo nos quedaremos con la siguiente selección de métodos: *SIFT*, *SURF*, *ORB*, *BRISK*, *AKAZE* y dos redes neuronales *SUPERPOINT* y *D2NET*, debido a que nuestro objetivo es evaluar métodos que han mostrado muy buen comportamiento en la extracción de características y métodos más recientes basados en aprendizaje profundo, como las redes neuronales.

A continuación, se va a describir con más detalle el funcionamiento de los métodos de descripción utilizados en este trabajo.

- *SIFT (Scale Invariant Feature Transform)*: Es un algoritmo de detección y de descripción. Como método de detección, detecta los

puntos mediante la función DoG (Función de diferencia de Gaussianas) aplicada en el espacio de escalas [13] .

Los puntos se seleccionan como extremos locales de la función DoG. A continuación, se calcula un descriptor para cada punto detectado, basado en la información local en la escala característica. Inicialmente fue utilizado en tareas de reconocimiento de objetos. Más tarde, ha sido aplicado a tareas de SLAM visual.

Las características SIFT son invariantes a escala, orientación y distorsión afín, y parcialmente invariantes a cambios de iluminación. La primera etapa, detección de los máximos y mínimos en el espacio de escalas consiste en la extracción de puntos característicos de la imagen y sus pasos son:

- a) Se realiza la convolución entre la imagen original $I(x, y)$ y una máscara Gaussiana $G(x, y, \sigma)$. Este resultado se guarda en una imagen A.

$$A(x, t, \sigma) = G(x, y, \sigma) \otimes I(x, y), \quad (1)$$

Donde
$$G(x, y, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x^2 + y^2)}{2\sigma^2}} \quad (2)$$

- b) A continuación, vuelve a repetir el proceso de convolución y el resultado se guarda en la imagen B

- c) Se calcula la diferencia de gaussianas (DoG): $C = A - B$.

- d) Las imágenes C y A se guardan para extraer posteriormente características SIFT.

- e) La imagen B se muestrea a diferentes escalas, de ese modo se obtiene una pirámide de imágenes DoG filtradas a diferentes escalas.

- f) Finalmente, en las imágenes DoG se buscan los máximos y los mínimos en un entorno local. De este modo se obtienen características SIFT a varias escalas de la imagen.

El proceso anteriormente descrito, se repite de forma sucesiva hasta que con el escalado, la imagen alcanza un valor tan pequeño que no permite extraer información útil.

En cuanto a método de descripción, este algoritmo se asigna una orientación global a cada punto basada en las direcciones del gradiente en un entorno local. Estas orientaciones se calculan de forma relativa a la orientación global de las características SIFT, consiguiendo la invarianza a la rotación. Por otra parte, los gradientes calculados se ponderan mediante una Gaussiana, de manera que los píxeles centrales reciben un mayor peso que el resto. A continuación se calcula un descriptor basado en los histogramas de orientación en una subregión 4x4 entorno al punto de interés, resultando un vector de dimensión 128 [13]. Para obtener invarianza a iluminación, que es lo que nos atañe en nuestros experimentos, el descriptor se normaliza con la raíz cuadrada de la suma de los componentes al cuadrado. En la Figura 8, podemos observar un ejemplo de este cálculo.

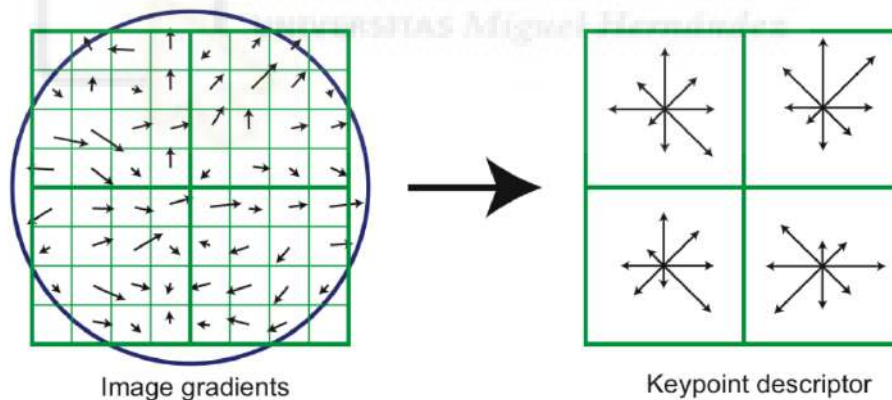


Figura 8. Descriptores SIFT.

- **SURF (Speeded Up Robust Features):** Similar al algoritmo SIFT, detecta y describe [14]. Como detector se basa en la matriz Hessiana y utiliza imágenes integrales para reducir el coste computacional. El detector SURF utiliza el determinante de la Hessiana para calcular la escala y la localización. Dado un punto $x = (x, y)$ en una imagen I , la matriz Hessiana $H(x, \sigma)$ en x a escala σ se define como sigue:

$$H(x, \sigma) = [L_{xx}(x, \sigma) L_{xy}(x, \sigma) L_{xy}(x, \sigma) L_{yy}(x, \sigma)] \quad (3)$$

Donde $L_{xx}(x, \sigma)$ es la convolución de la derivada Gaussiana de segundo orden $\frac{\delta^2}{\delta x^2} g(\sigma)$ con la imagen I en el punto x, y de forma similar para $L_{xy}(x, \sigma)$ y $L_{yy}(x, \sigma)$.

De forma similar al algoritmo del detector SIFT, SURF implementa el espacio de escalas con una pirámide de imágenes. Esta pirámide se obtiene filtrando las imágenes con un filtro Gaussiano y sucesivamente, realizando un muestreo de la imagen para alcanzar un nivel superior en la pirámide.

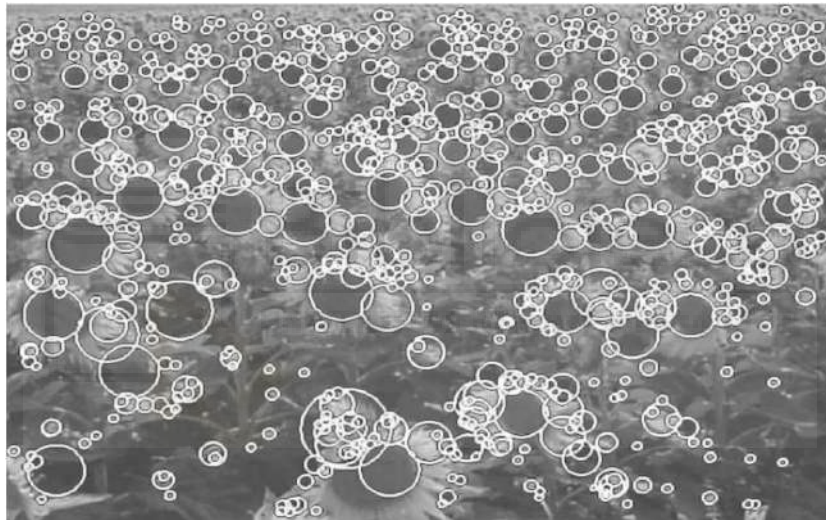


Figura 9. Puntos SURF.

En cuanto a método de descripción, representa una distribución de respuestas Haar-wavelet en el entorno del punto de interés y utiliza imágenes integrales eficientemente [14]. Existen tres versiones de este descriptor: SURF (estándar) de dimensión 64, la versión extendida (E-SURF) con 128 elementos y la versión *upright* (U-SURF), de 64 elementos. El descriptor SURF consiste en dos partes:

- 1) Asignación de la orientación. Para ello se calculan respuestas del filtro *wavelet*, con la transformada *Haar* en las direcciones x e y . Esto se realiza en una región circular de radio $6s$ alrededor del punto de interés, siendo s la escala a la cual el punto fue

detectado. Una vez que las respuestas del filtro wavelet han sido calculadas, y ponderadas con una Gaussiana ($\sigma = 2,5s$) centrada en el punto de interés, las respuestas se representan como vectores, donde la respuesta vertical corresponde al eje de ordenadas y la horizontal al eje de abscisas. A continuación, se calcula la orientación dominante a partir de la suma de todas las respuestas en una ventana de orientación corregida que cubre un ángulo de 60° . Las respuestas horizontal y vertical situadas en la ventana se suman, dando lugar a un nuevo vector. El vector de mayor tamaño será el que de su orientación al punto de interés.

- 2) Cálculo del descriptor. En primer lugar, hay que establecer una región cuadrada con centro en el punto de interés y orientada según la orientación seleccionada del paso anterior. La región cuadrada se divide de forma regular en 4×4 subregiones cuadradas más pequeñas. Para cada subregión, se calculan características en 5×5 puntos de muestreo espaciados regularmente. Se denota como dx a la respuesta *Haar wavelet* en la dirección horizontal y dy en la dirección vertical. Los términos *horizontal* y *vertical* están referidos a la orientación seleccionada para el punto de interés. A continuación, las respuestas del filtro dx y dy se suman para cada subregión, formando un primer vector de características. Además, se calcula la suma de los valores absolutos de estas respuestas ($|dx| + |dy|$). De este modo, para cada subregión se extrae un vector de 4 dimensiones $v = (\sum dx, \sum dy, \sum |dx|, \sum |dy|)$. Dado que la región se ha dividido en 4×4 subregiones, cada una con un vector de 4 elementos, el descriptor SURF estará formado por 64 elementos.

En la Figura 10, podemos ver un ejemplo de varias imágenes en el cálculo del descriptor SURF

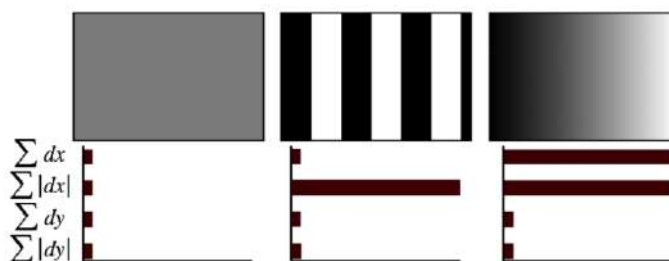


Figura 10. Descriptor SURF.

- **ORB, Oriented FAST and Rotated BRIEF:** Es también un detector y descriptor bastante eficiente que combina los algoritmos FAST, que detecta puntos característicos de manera rápida y BRIEF, para describir las características [23].

El detector FAST, comienza detectando puntos FAST en la imagen. Utiliza un parámetro, el umbral de intensidad entre el píxel central y aquellos en un anillo circular alrededor del centro. No produce una medida de "esquinidad" (*cornerness*), y tiene grandes respuestas a lo largo de los bordes. Emplea una medida de esquina de Harris, para ordenar los puntos clave de FAST. Para un número objetivo N de puntos clave, primero ajustamos el umbral lo suficientemente bajo como para obtener más de N puntos clave, luego los ordenamos según la medida de Harris y seleccionamos los N puntos principales.

No produce características en múltiples escalas. Utilizamos una pirámide de escalas de la imagen y producimos características FAST (filtradas por Harris) en cada nivel de la pirámide.

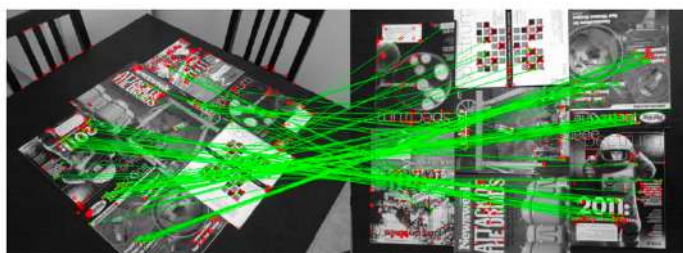


Figura 11. Resultado de coincidencias utilizando ORB.

El descriptor BRIEF, es un descriptor de características reciente que utiliza pruebas binarias simples entre píxeles en un parche de imagen suavizada. Su rendimiento es similar al de SIFT en muchos aspectos incluyendo robustez ante la iluminación, desenfoque y distorsión perspectiva.

El descriptor es una descripción en cadena de bits de un parche de imagen construido a partir de un conjunto de pruebas binarias de intensidad. Considera un parche de imagen suavizado, p . Una prueba binaria τ se define por (4):

$$\tau(p; x, y) := \begin{cases} 1 & : p(x) < p(y), \\ 0 & : p(x) \geq p(y) \end{cases} \quad (4)$$

donde ($p(x)$) es la intensidad de p en un punto x . La característica se define como un vector de n pruebas binarias (5):

$$f_n(p) := \sum_{1 \leq i \leq n} 2^{i-1} \tau(p; x_i, y_i) \quad (5)$$

Usamos una distribución Gaussiana alrededor del centro del parche y una longitud de vector ($n = 256$). Es importante suavizar la imagen antes de realizar las pruebas. En nuestra implementación, el suavizado se logra utilizando una imagen integral, donde cada punto de prueba es una subventana de 5×5 de un parche de 31×31 píxeles.

- **BRISK, Binary Robust Invariant Scalable Keypoints:** Este también se trata de otro método de detección y descripción. La detección de puntos característicos la realiza utilizando un enfoque de escala múltiple mediante pirámides de imágenes e intersección de escalas [25]. También utiliza un patrón de muestreo en anillo para extraer características alrededor del punto. Para cada punto, genera un descriptor binario comparando las intensidades de los puntos de muestreo en el patrón de anillo. Este método es invariante a la rotación y a la escala.

Detecta esquinas utilizando el algoritmo *AGAST* y las filtra con la puntuación *FAST Corner* mientras buscan máximos en la pirámide espacial de escala. La descripción de *BRISK* se basa en la identificación de la dirección característica de cada característica para lograr la invariancia de rotación. Para satisfacer la invarianza de la iluminación, los resultados de las pruebas de brillo simples también se concatenan y el descriptor se construye como una cadena binaria. Las características de *BRISK* son invariantes a la escala, la rotación y los cambios afines limitados.

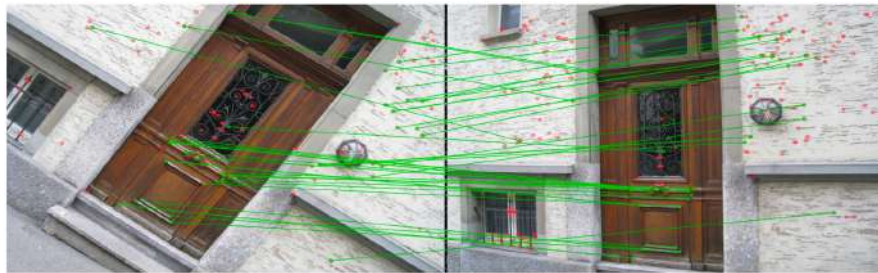


Figura 12. Resultado tras utilizar *BRISK*.

- **AKAZE:** Es otro método de descripción y detección de puntos característicos. Este algoritmo también sirve para realizar emparejamiento de puntos clave en imágenes [24]. Se basa en los conceptos de la Transformada No Lineal de Escala Espacial Acelerada (*A-KAZE*). En este caso, los puntos característicos se detectan empleando una la respuesta Hessiana, y genera descriptores basados en modificaciones de intensidad o descriptores binarios. La calidad de la invariancia de rotación se mejora utilizando *los filtros Scharr*. Las máximas de las respuestas del detector en ubicaciones espaciales se recogen como puntos característicos. El descriptor de *AKAZE* se basa en el algoritmo *modificado de diferencia local binaria (MLDB)* que también es altamente eficiente. Las características de *AKAZE* son invariantes a la escala, la rotación, la afinidad limitada y tienen más carácter distintivo en diferentes escalas debido a los espacios de escala no lineales.



Figura 13. Coincidencias utilizando AKAZE.

A) REDES NEURONALES

Una red neuronal es un sistema de *machine learning* que toma decisiones de manera similar al cerebro humano, utilizando procesos que imitan cómo las neuronas biológicas trabajan juntas para identificar patrones, evaluar opciones y llegar a conclusiones [19].

Una red neuronal está compuesta por capas de nodos o neuronas artificiales: una capa de entrada, una o varias capas ocultas y una capa de salida. Cada nodo está conectado a los demás y tiene sus propias ponderaciones y umbrales. Si la salida de un nodo supera el umbral especificado, ese nodo se activa y envía información a la siguiente capa de la red; si no, no envía nada.

Las redes neuronales aprenden y mejoran su precisión con el tiempo usando datos de entrenamiento. Una vez entrenadas, son herramientas poderosas en informática e inteligencia artificial, capaces de clasificar y agrupar datos rápidamente. Tareas como el reconocimiento de voz o de imágenes pueden realizarse en minutos, en comparación con las horas que tomarían si fueran hechas manualmente por expertos humanos. Un ejemplo bien conocido de red neuronal es el algoritmo de búsqueda de Google.

Las redes neuronales también se conocen como redes neuronales artificiales (ANN) o redes neuronales simuladas (SNN). Son una

subcategoría del *machine learning* y forman la base de los modelos de *deep learning*.

Hay varios tipos de redes, como:

- *Redes neuronales de avance o perceptrones multicapa (MLP)*. Estas redes consisten en una capa de entrada, una o más capas ocultas y una capa de salida. Aunque comúnmente se llaman MLP, es importante destacar que están formadas por neuronas sigmoideas en lugar de perceptrones, ya que la mayoría de los problemas del mundo real son no lineales. Estos modelos se entrenan introduciendo datos, y son fundamentales para la visión por computadora, el procesamiento del lenguaje natural y otras aplicaciones de redes neuronales.
- *Las redes neuronales convolucionales (CNN)* son similares a las redes de avance (*feedforward*), pero se utilizan principalmente para el reconocimiento de imágenes, el reconocimiento de patrones y la visión por computadora. Estas redes aplican principios del álgebra lineal, especialmente la multiplicación de matrices, para identificar patrones en las imágenes.
- *Las redes neuronales recurrentes (RNN)* se caracterizan por sus bucles de retroalimentación. Estos algoritmos de aprendizaje se usan principalmente con datos de series temporales para hacer predicciones sobre futuros resultados, como en las predicciones bursátiles o las previsiones de ventas.

Hoy en día, las redes neuronales se utilizan en infinidad de aplicaciones, como por ejemplo en reconocimiento facial, asistentes virtuales como Siri o Alexa. En seguridad, el reconocimiento biométrico. Por otra parte en automóviles y control, en sensores para la navegación autónoma. También se utilizan en el ámbito de la medicina, para diagnósticos con imágenes o predicción de enfermedades.

Nosotros nos vamos a centrar en el uso de las redes neuronales para la detección y descripción del entorno. Estas redes son capaces de procesar grandes cantidades de datos visuales y sensoriales para identificar objetos o mapear entornos. Un ejemplo de aplicación de este tipo de redes es el Autopilot de los coches Tesla [20]. Esta función tan conocida se basa en un sistema de 8 cámaras externas equipadas con visión por ordenador (Figura 14). Tesla, utiliza la información que captan con las cámaras todos los coches Tesla, para crear una base de datos inmensa para entrenar las redes.

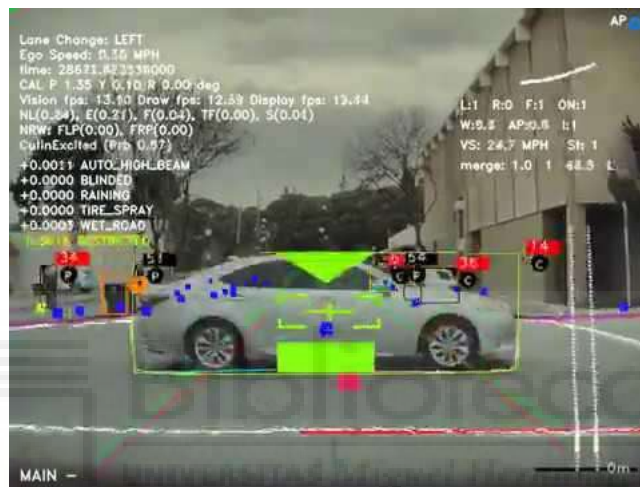


Figura 14. Imagen que “observa” la red neuronal en Autopilot de Tesla.

- **SUPERPOINT:**

Se trata de una red neuronal convolucional que actúa como detector y como descriptor. Esta red, opera con imágenes de tamaño completo y calcula conjuntamente las ubicaciones de los puntos de interés a nivel de cada píxel y los descriptores asociados en una sola pasada [21].

Los puntos de interés son ubicaciones 2D en una imagen que son estables y repetibles bajo diferentes condiciones de iluminación y puntos de vista. En lugar de utilizar la supervisión humana para definir puntos de interés en imágenes reales, propone una solución auto-supervisada utilizando auto entrenamiento. Por lo que crea un gran conjunto de datos

de ubicaciones de puntos de interés de pseudo-verdad en imágenes reales, supervisado por el propio detector de puntos de interés, en lugar de un trabajo por parte nuestra para revisarlo.

Para generar estos puntos, entrenamos una red neuronal convolucional en millones de ejemplos de un conjunto de datos sintéticos que creamos llamado Formas Sintéticas. Llamamos al detector entrenado resultante denominado *MagicPoint*. Este detector, supera con creces a los detectores de puntos tradicionales en el conjunto de datos sintético. *MagicPoint* funciona muy bien en imágenes reales a pesar de las dificultades de adaptación al dominio, pero en cambio, a diferencia de los detectores clásicos en un conjunto diverso de texturas y patrones de imágenes, pierde muchas posibles ubicaciones de puntos de interés. Para solucionar esto, surge una técnica multi-escala y multi-transformación llamada Adaptación Homográfica.

La Adaptación Homográfica está diseñada para permitir el entrenamiento auto supervisado de detectores de puntos de interés. Deforma la imagen de entrada múltiples veces para ayudar a un detector a ver la escena desde muchos puntos de vista y escalas diferentes. Se utiliza la Adaptación Homográfica junto con el detector *MagicPoint* para mejorar el rendimiento del detector y generar los puntos de interés de pseudo-verdad. Las detecciones resultantes son más repetibles y responden a un conjunto más amplio de estímulos, y a este detector resultante se le conoce como *SuperPoint*.

Lo más común después de detectar puntos de interés robustos y repetibles, es adjuntar un vector descriptor de dimensión fija a cada punto para tareas semánticas de nivel superior, como emparejamiento de imágenes.

Por otra parte, en cuanto a la arquitectura de *SuperPoint*, el modelo tiene un codificador compartido único para procesar y reducir la dimensionalidad de la imagen de entrada. Después del codificador, la arquitectura se divide en dos decodificadores, que aprenden pesos

específicos de la tarea: una para la detección de puntos de interés y la otra para la descripción de puntos de interés.

En resumen, podemos decir que la red neuronal *SuperPoint*, es un método de detección y descripción de puntos de interés el cual es muy robusto y repetible, gracias al entrenamiento auto-supervisado, el cual se realiza en dos etapas, *MagicPoint* y la Adaptación Homográfica.

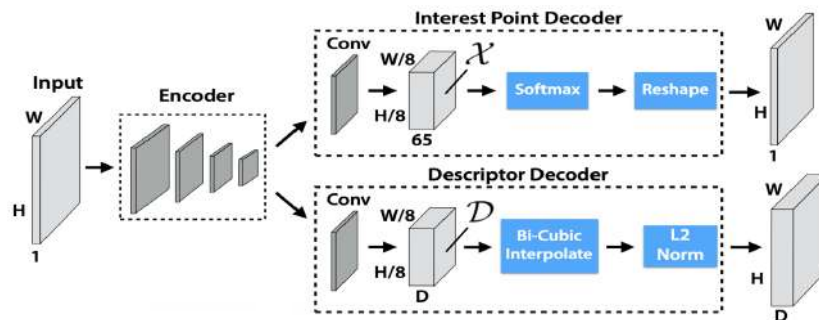


Figura 15. Estructura de capas de la red *SuperPoint* [21].

- **D2NET:**

Es una red neuronal completamente convolucional que se utiliza como método de detección y descripción simultáneamente, es decir que realiza la detección y descripción de puntos característicos de forma conjunta, a diferencia de los métodos clásicos, que realizan la detección y descripción en dos etapas [22]. En este caso se realiza una extracción de características densa para obtener una representación que simultáneamente sea un detector y un descriptor. Este enfoque se denomina D2, y mejora la repetibilidad y distintividad de los puntos bajo cambios de iluminación y vista.

La detección de puntos de interés es un paso crucial en muchas tareas de visión por computadora, como SLAM, calibración de cámaras y emparejamiento de imágenes. Tradicionalmente estos puntos se detectan con métodos como SIFT o FAST y luego se describen utilizando

descriptores de características. Sin embargo, esto tiene limitaciones cuando se enfrentan a variaciones significativas en las condiciones de iluminación o puntos de vista.

Al contrario de los enfoques clásicos de detección-descripción, se propone realizar la extracción de características de manera densa para obtener una representación que simultáneamente sea un detector y un descriptor. Esta representación compartida, denominada D2, se obtiene aplicando una CNN (Red neuronal convolucional) a la imagen de entrada, resultando un tensor en 3D. Este tensor en 3D se interpreta como un conjunto denso de vectores descriptor. Estos vectores se comparan entre imágenes utilizando la distancia euclidiana. Durante la fase de entrenamiento, estos descriptores se ajustan para que los mismos puntos en la escena produzcan descriptores similares, incluso cuando las imágenes contienen cambios de apariencia fuertes. Otra interpretación del tensor 3D es una colección de respuestas 2D. Estas respuestas son análogas a los mapas de respuesta de la diferencia de gaussianas (DoG) obtenidas en SIFT o a los mapas de puntuación de esquinas obtenidos en el detector de esquinas de Harris.

Para entrenar el modelo propuesto, que utiliza una única CNN tanto para la detección como para la descripción, se requiere una pérdida apropiada que optimice conjuntamente los objetivos de detección y descripción.

En resumen, la red D2NET se puede considerar un avance considerado en la detección y descripción de forma conjunta, ya que optimiza de manera eficiente y robusta estas dos tareas.

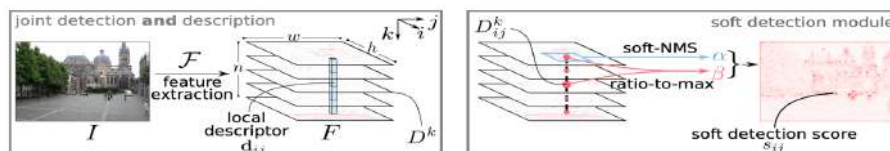


Figura 16. Estructura de capas de la red D2-NET [22].

3.2.2. MÉTODOS DE CARACTERÍSTICAS GLOBALES

Como hemos descrito anteriormente, la presencia de los robots móviles en diversos ámbitos han incrementado notablemente (ámbitos industriales, domésticos...), pero este despegue no será definitivo hasta que los robots no sean completamente autónomos y se adapten perfectamente a las diferentes condiciones del entorno. El robot autónomo debe poder encontrar solución rápidamente a los dos problemas comentados en un tiempo aceptable, la construcción de mapas (*mapping*) y la estimación de la posición en el mapa (localización). Si encuentra solución para estos dos problemas, realiza la tarea de SLAM.

Para ello, se emplean los sensores, para capturar la información del entorno, principalmente cámaras, ya que son de utilidad para las tareas de navegación. Las dos formas que hay de extraer información de una imagen son la extracción de puntos característicos que hemos estado viendo, o el tratamiento de la imagen en su conjunto, sin extraer características locales. Esta segunda será la que estudiemos en el presente apartado.

Esta técnica de tratar la imagen de forma global puede dar buenos resultados, como simplificar la estructura de los mapas, además de reducir el proceso de localización, ya que únicamente habrá que comparar descriptores globales. Por otra parte, una desventaja que tiene es que, al trabajar con una gran cantidad de datos, es necesario usar una técnica de compresión que reduzca el coste computacional. De aquí surgen los descriptores de apariencia global, que son métodos que reducen la información, de manera que también se reduce el coste computacional, pero manteniendo información necesaria para las tareas de *mapping* y localización.

En tareas realizadas en un entorno real, el robot se encuentra con diferentes retos, como los cambios en las condiciones de iluminación, escenas ocluidas por personas u otros robots móviles y cambios en la escena, como movimientos del mobiliario, por ejemplo. Nosotros evaluaremos la fiabilidad de algunos descriptores, en base a cambios de iluminación también.

A continuación, describiremos algunos descriptores de apariencia global, los cuales extraen información partiendo de imágenes panorámicas, y transformándolas mediante los procesos de descripción global a vectores [26]. Existen otros métodos como *GIST*, introducido en [35], y utilizado en otro estudio en el que se probó en entornos de exterior [32]. O la *Transformada de Radon*, descrita en [33], y utilizada en tareas de visión por computador como puede ser la descripción de formas geométricas, o en tareas del campo de la robótica donde fue utilizada para encontrar un vecino más cercano en un estimador de posición [36].

En este trabajo, estudiaremos *Firma de Fourier* y el método *Histograma de orientación del Gradiente*. Ya que ofrecen resultados más rápidos y son opciones adecuadas para tareas en tiempo real [26].

- **FIRMA DE FOURIER:** este descriptor calcula la transformada discreta de Fourier para cada fila de la imagen panorámica $i(x,y)$ y se obtiene una matriz compleja $F(u,v)$, la cual se puede dividir en dos matrices : una de módulos $A(u,v)$ una de argumentos $\angle(u,v)$. Dos imágenes que hayan sido capturadas desde la misma posición, pero con distintas orientaciones, tendrán la misma matriz de módulos, pero matrices de argumentos distintas. Por lo que la matriz A , se puede usar como descriptor de posición y la matriz \angle , se puede usar como descriptor de orientación. Por otra parte, en el dominio de Fourier, la información más significativa se concentra en las frecuencias bajas, por lo que se pueden eliminar las frecuencias altas, donde se almacena ruido. Esto permite minimizar el coste computacional. Se usan dos parámetros relevantes, k_1 , el cual índice el número de frecuencias retenidas, y N_x que indica el número de filas que tiene inicialmente la imagen panorámica. De esta manera, la imagen queda reducida a un descriptor $\vec{d} \in R^{N_x * k_1 * 1}$.

Hay autores como [27] que estudian este método para imágenes panorámicas. Y otros como [28] que proponen aplicar la transformada de Fourier esférica (SFT) a la imagen omnidireccional.

En la Figura 17, podemos observar un ejemplo del costo computacional de este método [29]

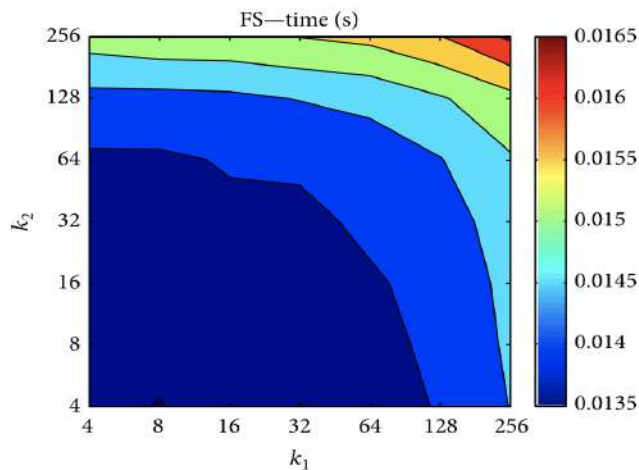
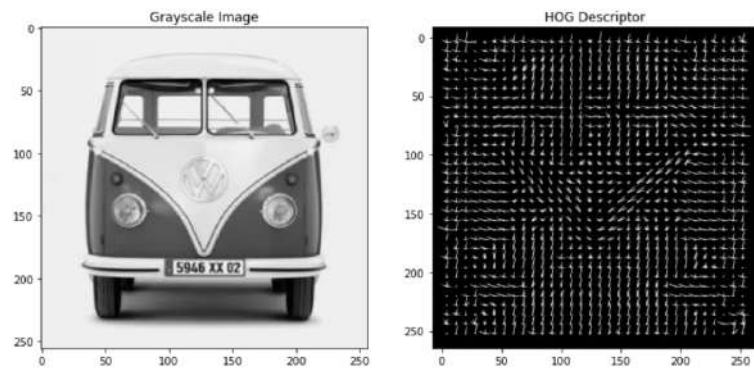


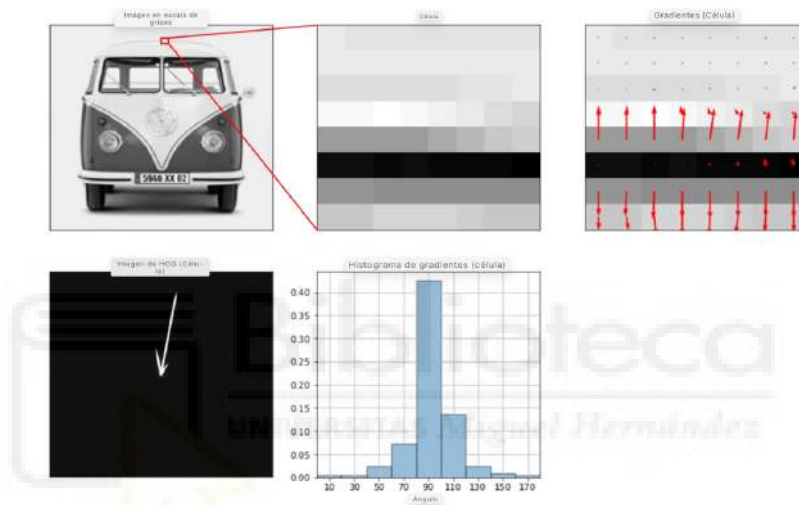
Figura 17. Costo computacional del método Firma de Fourier.

- HISTOGRAMA DE ORIENTACIÓN DEL GRADIENTE (HOG) :** Esta técnica consiste en calcular el gradiente de una imagen y calcular orientación y módulo de cada píxel [31]. A continuación, la imagen se divide en un conjunto de celdas horizontales, y se crean b_1 recipientes denominados *bins*. Dependiendo de la orientación obtenida antes y por celdas horizontales, cada píxel, ponderado por su valor de magnitud de gradiente, se almacenará en un *bin* distinto. De esta manera, dos imágenes panorámicas capturadas en el mismo punto tendrán la misma información en cada celda horizontal. Dando resultado a un descriptor invariante a la rotación del robot en el plano del suelo y puede ser utilizado como un descriptor de apariencia global. Además, el vector obtenido tiene un tamaño reducido que dependerá del número de celdas y *bins*. Siendo b_1 el número de *bins* y k_2 el número de celdas horizontales. El método reduce la información a un descriptor $\vec{d} \in R^{b_1 * k_2 * 1}$.

En la Figura 18.a podemos observar un ejemplo de una imagen en la que se aplica este método [30]. En la Figura 18.b, seleccionamos una celda en la imagen que encierra un borde horizontal. Los bordes son áreas de una imagen donde la densidad de píxeles cambia abruptamente. Cuando examinamos los gradientes, vemos una alta tasa de cambios verticales. Por esta razón vemos en el histograma, el contenedor de 90 grados será muy dominante en comparación con los otros.



(a)



(b)

Figura 18. En la figura 18.a. tenemos la imagen HOG. En la figura 18.b. tenemos las celdas y el histograma HOG.

Los histogramas de orientación se calculan a partir de imágenes gradientes. Para cada píxel, se calcula un módulo y una orientación. Los valores de orientación se dividen en intervalos y el histograma se forma con los valores del módulo. Estos se aplican en navegación de robots móviles [17]. Un histograma es una representación gráfica de barras, donde en el eje de abscisas se representan los valores de una variable y en el eje de ordenadas se representa la frecuencia de aparición de dichos valores.

Para construir un histograma de orientación, se calcula primero el gradiente de la imagen en la dirección x (dx) y en la dirección y (dy). De este modo, para cada píxel de la imagen, se calcula como en la ecuación (6a):

$$valor_{pixel} = atan(dx, dy), \quad (6a)$$

Y el módulo como (6b):

$$Módulo = \sqrt{dx^2 + dy^2} \quad (6b)$$

A continuación, se divide el rango de valores de orientación en partes iguales (*bins*). Dependiendo del número de *bins* elegido, el descriptor tendrá un tamaño determinado. Finalmente se calcula el módulo medio de toda la imagen y se establece un umbral igual a k veces dicho módulo. Para cada píxel, si el valor de su módulo supera este umbral, se aumentará la frecuencia de aparición correspondiente al valor del argumento de dicho punto.

4 ESTUDIO DE DESCRIPTORES VISUALES ANTE CAMBIOS DE ILUMINACIÓN

Hasta ahora hemos descrito de manera detallada los métodos de detección y descripción propuestos por diversos autores, que tienen como objetivo encontrar los puntos característicos que mejor funcionen en SLAM visual.

Como hemos comentado anteriormente, en situaciones reales las imágenes que se procesan en los diferentes métodos sufren variaciones, como pueden ser: cambios de iluminación, cambios de escala y de punto de vista, entornos interiores y exteriores.

4.1 – OBJETIVOS

En el presente trabajo nos centraremos en la evaluación de descriptores tanto locales como globales, pero únicamente ante cambios de iluminación. Para llevar

a cabo esta tarea utilizaremos dos bases de datos, “BASLER” y “ARVC” consistentes en secuencias de imágenes con cambios de iluminación, donde habrá secuencias de entornos exteriores e interiores. Podemos ver un ejemplo de dos secuencias de las distintas bases de datos en la figura 19a) y 19b).



a)



b)

19. La secuencia a) es la número 1 de la base de datos BASLER-exterior. Y la secuencia b) es la número 2 de la base de datos ARVC de interior.

El estudio consistirá en procesar las imágenes mediante algunos métodos de descripción y una vez obtenidos los descriptores de cada secuencia de imágenes, utilizaremos una serie de métricas que definiremos a continuación para poder evaluar y comparar los métodos utilizados.

4.2 – MÉTRICAS PARA EVALUAR LOS EXPERIMENTOS REALIZADOS

En primer lugar, emplearemos una métrica de evaluación para los detectores, Ratio de repetibilidad, y a continuación otras dos para los descriptores, Curvas *recall vs precision* y Criterio de separabilidad. Para los descriptores locales, emplearemos estas tres métricas, pero para los descriptores globales, únicamente evaluaremos con las métricas, Curvas *recall vs precision* y Criterio de separabilidad [4]. Esto se debe a que, en este caso, cada imagen se describe

con un descriptor, por lo que no es necesario comprobar que los puntos detectados se encuentran en las mismas posiciones a lo largo de la secuencia.

- *Ratio de repetibilidad*: Esta métrica evalúa el detector y comprueba que los puntos de interés que detecta, siempre se encuentren en las mismas posiciones, aunque la iluminación varíe [9]. Tras realizar el seguimiento de los puntos visto en el apartado 3.2, se podrá definir una ratio de supervivencia rs_i en la imagen i de la secuencia como en la ecuación (7):

$$rs_i = \frac{np_i}{np_1} * 100 \quad (7)$$

Donde np_i es el número de puntos de interés seguidos hasta la imagen i en la secuencia y np_1 es el número de puntos de interés detectados en la primera imagen de la secuencia.

El detector ideal sería el que detectase los mismos puntos a lo largo de la secuencia ($rs_i = 100\%$) para cada imagen de la secuencia. Pero lo usual suele ser que haya una tendencia decreciente de rs_i , lo que significa que algunos de los puntos observados en la primera imagen se pierden en las siguientes imágenes de la secuencia.

Para conseguir marcas visuales que sean robustas y estables, se hace un seguimiento de cada punto a lo largo de todas las imágenes de la secuencia y se incluyen únicamente los puntos que se hayan detectado en p imágenes consecutivas. Como resultado, el número de marcas se reduce y por tanto también la complejidad del problema SLAM. Para considerar este requerimiento se analiza durante cuántas imágenes se debe seguir una marca antes de integrarla en el mapa, y se utiliza la siguiente probabilidad condicionada (8)

$$P(I_i) = \frac{n_{p1,j}}{n_{p1,i}}, \quad j \geq i \quad (8)$$

Donde, $n_{p1,j}$ es el número de puntos que puede ser seguido desde la primera hasta la imagen k en la secuencia. Este valor representa la probabilidad de que un punto de interés se siga hasta la imagen

I_j dado que se siguió hasta la imagen I_i este valor oscila entre 0 y 1. Toma el valor de 0 cuando todos los puntos seguidos hasta la imagen I_i se pierden en la imagen I_j y 1 si ambas imágenes contienen los mismos puntos seguidos. Este valor interesante cuando I_j es la última imagen de la secuencia. En este caso $P(I_i)$ nos da la probabilidad de que un punto detectado esté en la última imagen.

La expresión (8) proporciona una predicción de la supervivencia de un punto de interés en imágenes consecutivas si el movimiento del robot se mantiene similar. Esta expresión se puede utilizar para estimar el número p de imágenes a lo largo de las cuales debe seguirse un punto antes de ser incorporado en el mapa.

- *Curvas de Recall vs Precision:* Mediante esta métrica se evalúa la cantidad de asociaciones correctas realizadas y el coste relativo de los falsos positivos [9]. Como hemos visto, la evaluación se realiza considerando que cada marca visual se representa mediante un *cluster*, formado por descriptores del mismo punto obtenidos desde diferentes puntos de vista, ya que esta situación puede darse frecuentemente en SLAM visual.

Mediante el método de seguimiento explicado en el apartado 3.2, un punto de interés x , que haya sido encontrado en al menos dos imágenes en una secuencia, será representado por M conjuntos diferentes D_1^x, \dots, D_M^x . De este modo, el conjunto $D_M^x = \{d_m^x(1), \dots, d_m^x(N)\}$ se considera como un *cluster* y representa el punto a lo largo de la trayectoria utilizando el método de descripción m , siendo N el número de imágenes donde fue encontrado el punto. Dado un descriptor que tiene una vista particular de una marca visual, el objetivo es encontrar su *cluster* correspondiente, utilizando una medida de distancia, en este caso la distancia Euclídea (9) :

$$E = \sqrt{(d_m^{x_i} - d_m^{x_j})^T (d_m^{x_i} - d_m^{x_j})} \quad (9)$$

Donde $d_m^{x_i}$ es un descriptor que pertenece a la clase ω_i y $d_m^{x_j}$ es un descriptor asociado a la clase ω_j . Para cada descriptor $d_m^{x_i}$ que queremos clasificar, se calcula la distancia Euclídea a todos los *clusters* y se busca el *cluster* que minimice esta distancia. Las correspondencias correctas son conocidas para todos los descriptores: una correspondencia es verdadera cuando el descriptor $d_m^{x_i}$ se asigna al *cluster* ω_i y falsa cuando el descriptor se asigna a un *cluster* diferente. Como resultado se obtiene una lista de descriptores, cada uno con una distancia Euclídea mínima asociada a un *cluster* y una correspondencia verdadera/falsa. A continuación, se ordena la lista de asociaciones de forma ascendente en base a la distancia Euclídea mínima y se utiliza para calcular los parámetros *recall* y *precision*, que se definen como (10) y (11):

$$recall = \frac{\text{num.correspondencias correctas seleccionadas}}{\text{num.total correspondencias correctas}} \quad (10)$$

$$precision = \frac{\text{num.correspondencias correctas seleccionadas}}{\text{num. correspondencias seleccionadas}} \quad (11)$$

- Recall*: habilidad de encontrar todas las asociaciones correctas
- Precision*: capacidad de obtener asociaciones correctas cuando el número de asociaciones varía.
- Factor Num. Correspondencias seleccionadas*: representa el número de asociaciones en la lista cuya distancia es menor que el umbral establecido. Dicho factor puede variar entre 1 y el número total de asociaciones que componen la lista
- Variable Num. Correspondencias correctas seleccionadas*: es el número de correspondencias correctas obtenidas dado un determinado umbral en la lista ordenada.
- Factor Num. Total correspondencias correctas*: es un valor constante, que expresa el número total de correspondencias correctas en la lista.

En la curva *precision vs recall*, un valor alto de *precision* con un valor de *recall* bajo indica que se han obtenido asociaciones correctas, pero que se han perdido varias. Por otro lado, un alto valor de *recall* con un valor bajo de *precision* supone que se han encontrado la mayoría de las asociaciones correctas pero que también existen muchas incorrectas. Por lo que la situación ideal sería encontrar un descriptor que obtenga valores altos de ambos parámetros simultáneamente.

- *Criterio de separabilidad*: Las curvas *recall vs precision* representa el comportamiento de los descriptores en diferentes situaciones, pero es necesario una medida que permita establecer un ranking de descriptores en base a cuál es mejor en SLAM visual, por ello el criterio de separabilidad de *clusters* es una buena opción .

Para estudiar la separabilidad de los *clusters* que representan a los puntos de interés se utiliza el criterio de separabilidad J3 [34]. Esta medida se basa en dos matrices de dispersión: S_w y S_b . S_w es la matriz intra-clase y mide la varianza intra-clase de un *cluster*. Y S_b es la matriz inter-clase, que mide la distancia entre *clusters* diferentes.

Para un *clustering* dado, S_w se calcula como:

$$S_w = \sum_{i=1}^V P_i S_i \quad (12)$$

Donde S_i es la matriz de covarianza para la clase ω_i :

$$S_i = E[(x - \mu_i)(x - \mu_i)^T] \quad (13)$$

P_i es la probabilidad a priori de la clase ω_i y μ_i es el descriptor medio para la clase ω_i .

La matriz S_b se representa como:

$$S_b = \sum_{i=1}^V P_i (\mu_i - \mu_o)(\mu_i - \mu_o)^T \quad (14)$$

Donde μ_o es la media global calculada como:

$$\mu_o = \sum_{i=1}^V P_i \mu_i \quad (15)$$

$J3$ se define como:

$$J3 = \text{trace} (S_w^{-1} S_m). \quad (16)$$

donde S_m es la matriz de dispersión mixta, y se calcula como la suma de las anteriores:

$$S_m = S_w + S_b \quad (17)$$

Un buen descriptor será aquel que tenga un valor bajo de S_w , ya que indica la variabilidad de los vectores describiendo la misma clase debe ser pequeña. Además, es deseable que los vectores que describen diferentes puntos estén diferenciados del resto, resultando un valor alto de S_b . Como consecuencia, un buen descriptor tendrá un valor alto de $J3$. Con el fin de comparar descriptores que tengan longitudes diferentes se utiliza una versión normalizada de este criterio:

$$J'_3 = \frac{J3}{L} \quad (18)$$

Donde L es la longitud del descriptor.

4.3 – EXPERIMENTOS Y RESULTADOS

Hemos realizado el estudio diferenciando descriptores locales y globales, y como hemos comentado, todos los resultados los hemos obtenido para las bases de datos BASLER y ARVC, además de secuencias interiores y exteriores.

A continuación, mostraremos en las figuras 20 y 21, el número de secuencias que hay en cada base de datos y el número de imágenes que hay por cada secuencia.

BASE DATOS	Nº Secuencia Interior	Nº Imágenes por Secuencia
BASLER	1	28
ARVC	1	14
	2	9
	3	10

Figura 20. Tabla con el número de secuencias e imágenes interiores

BASE DATOS	Nº Secuencia Exterior	Nº Imágenes por Secuencia
BASLER	1	14
	2	21
ARVC	1	15
	2	15
	3	15
	4	16
	5	14
	6	17
	7	17

Figura 21. Tabla con el número de secuencias e imágenes exteriores

A continuación se muestran algunas de las imágenes correspondientes a cada una de las secuencias de imágenes utilizadas para posteriormente evaluar los descriptores en ellas.



Figura 22.. Imágenes 1,8,14 y 27 de la secuencia 1 de interior - BASLER



Figura 23. Imágenes 1,4,9 y 14 de la secuencia 1 de exterior - BASLER



Figura 24. Imágenes 1,6,12 y 20 de la secuencia 2 de exterior - BASLER



Figura 25.. Imágenes 1,4,9 y 13 de la secuencia 1 de interior- ARVC



Figura 26.. Imágenes 1,3,6 y 9 de la secuencia 2 de interior - ARVC



Figura 27. Imágenes 1,4,7 y 9 de la secuencia 3 de interior - ARVC



Figura 28. Imágenes 1,5,10 y 15 de la secuencia 1 de exterior - ARVC



Figura 29. Imágenes 1,4,11 y 15 de la secuencia 2 de exterior - ARVC



Figura 30. Imágenes 1,3,7 y 14 de la secuencia 3 de exterior - ARVC



Figura 31. Imágenes 1,7,11 y 16 de la secuencia 4 de exterior - ARVC



Figura 32. Imágenes 1,6,11 y 17 de la secuencia 6 de exterior- ARVC



Figura 33. Imágenes 1,4,8 y 15 de la secuencia 7 de exterior - ARVC

4.3.1. Métodos de características locales

Los métodos que evaluaremos serán: SIFT, SURF, ORB, BRISK, AKAZE, D2NET y SUPERPOINT, descritos en el apartado 3.2.1.

En primer lugar, en cada secuencia de imágenes calcularemos para cada método el número de puntos característicos y de *clusters* obtenidos, como podemos observar en la Figura 34, se trata de un ejemplo obtenido de una secuencia de imágenes de interior de la base de datos BASLER. Estos resultados también serán obtenidos para cada secuencia para obtener los gráficos correspondientes.

RED	Puntos característicos en la imagen de referencia (12)	Nº Clusters
D2NET	12984	17
SIFT	2886	47
ORB	7662	236
BRISK	2396	66
AKAZE	1639	91
SUPERPOINT	2738	307
SURF	5552	290

Figura 34. Puntos característicos y número de clusters obtenidos para una secuencia de interior (BASLER).

a) Ratio de repetibilidad

A continuación, se van a mostrar los gráficos obtenidos para el Ratio de Repetibilidad con cada descriptor. Para mostrar estos datos, se ha diferenciado entre los resultados obtenidos en las secuencias de la base de datos BASLER y las secuencias ARVC, a su vez, se ha diferenciado entre secuencias interiores y exteriores (Figuras 20 y 21).

Cada gráfica muestra, para cada descriptor, el valor de ratio de repetibilidad vs el brillo medio de la imagen. La imagen de referencia es la que ocupa una posición central en base a la magnitud de brillo. El objetivo, como se ha comentado anteriormente, es valorar cómo de estables son los puntos característicos detectados a pesar de los cambios de iluminación en la imagen.

- BASLER:
 - INTERIOR:

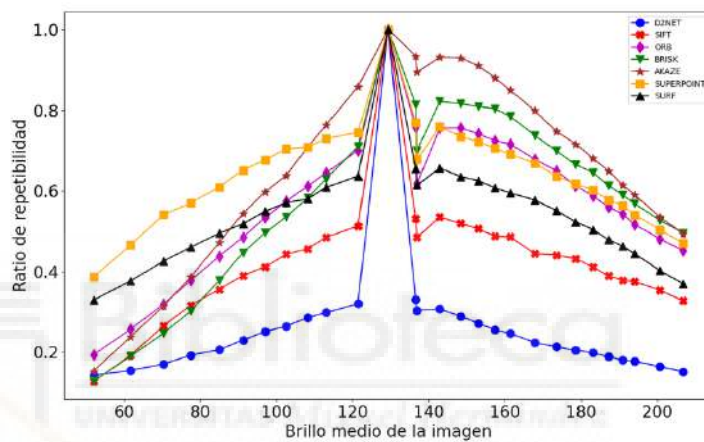


Figura 35. Ratio de repetibilidad - BASLER- database - indoor -sec1

- EXTERIOR

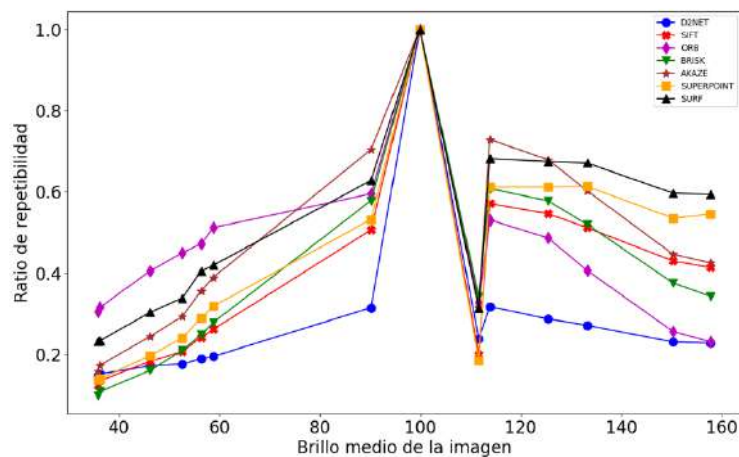


Figura 36. Ratio de repetibilidad - BASLER- database - outdoor -sec1

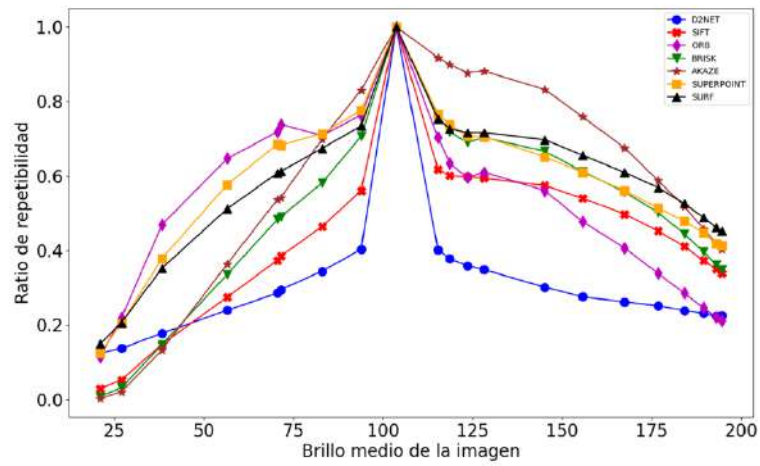


Figura 37. Ratio de repetibilidad - BASLER- database - outdoor -sec2

- ARVC:
 - INTERIOR

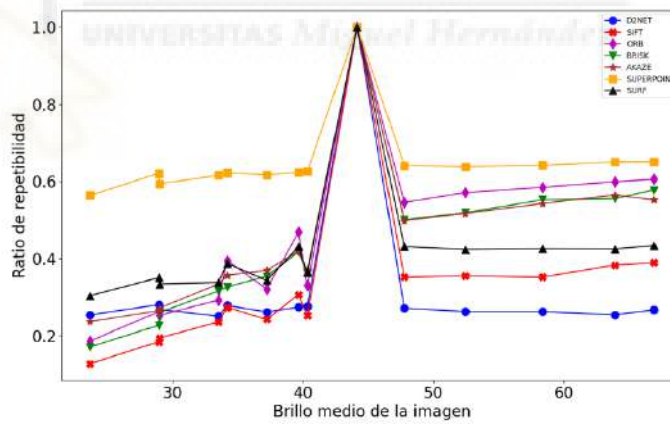


Figura 38. Ratio de repetibilidad - ARVC- database - indoor -sec1

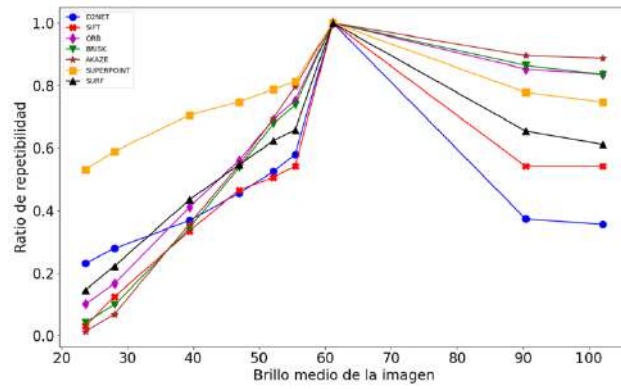


Figura 39. Ratio de repetibilidad - ARVC- database - indoor -sec2

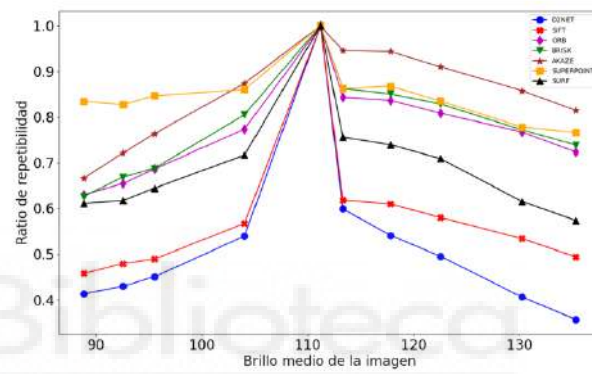


Figura 40. Ratio de repetibilidad - ARVC- database - indoor -sec3

○ EXTERIOR

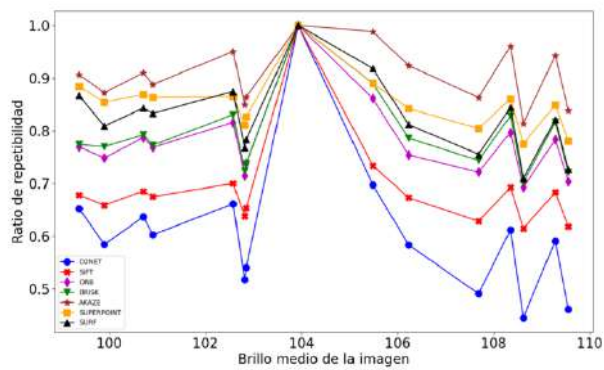


Figura 41. Ratio de repetibilidad - ARVC- database - outdoor -sec1

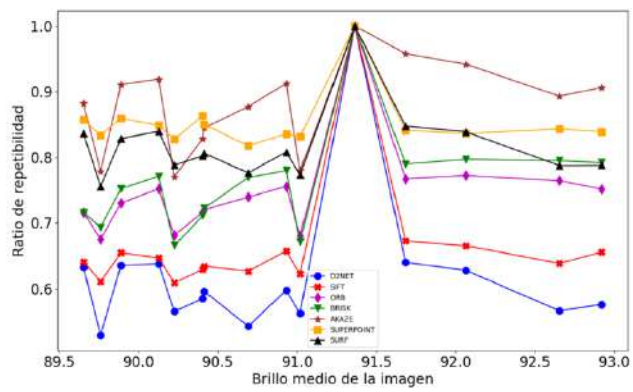


Figura 42. Ratio de repetibilidad - ARVC- database - outdoor -sec2

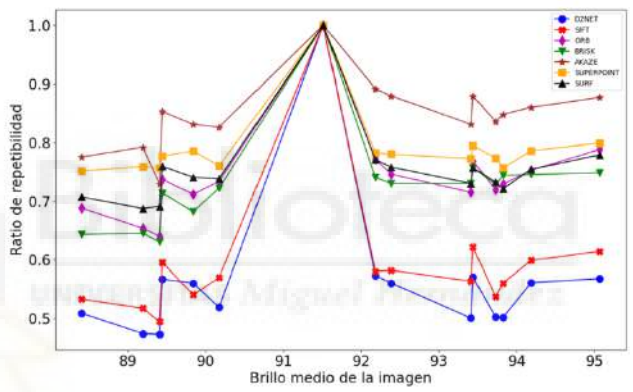


Figura 43. Ratio de repetibilidad - ARVC- database - outdoor -sec3

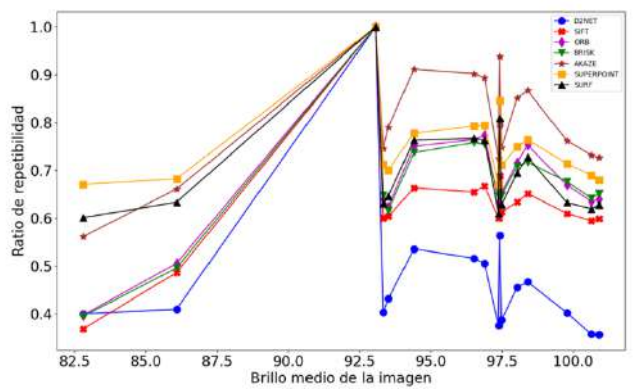


Figura 44. Ratio de repetibilidad - ARVC- database - outdoor -sec4

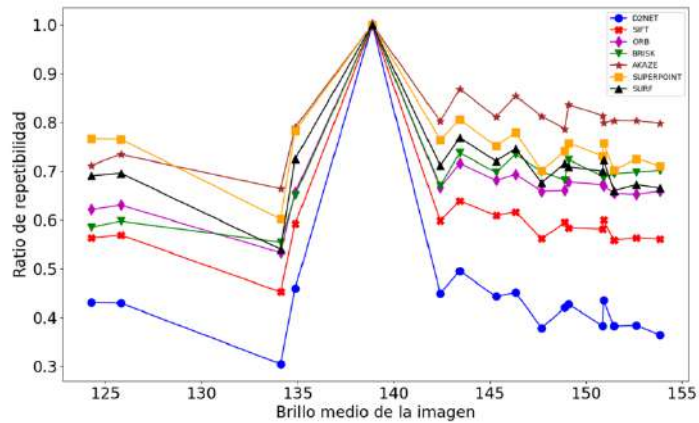


Figura 45. Ratio de repetibilidad - ARVC- database - outdoor -sec6

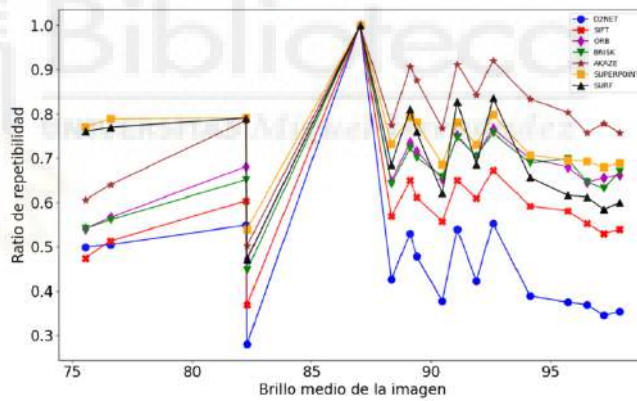


Figura 46. Ratio de repetibilidad - ARVC- database - outdoor -sec7

Como hemos comentado en el apartado 4.2, la situación ideal de un método sería el que se tenga $rs_i = 1$ y de manera estable. A raíz de los resultados obtenidos podemos comprobar que la detección de puntos característicos tiene mejores resultados en entornos de interior que de exterior. Esto podemos observarlo en las Figuras 35, y de la 38 a la 40, que son las secuencias de interior, y presentan unos valores de repetibilidad más estables. Por ejemplo, si nos fijamos en la figura 40, de interior, podemos observar que los métodos presentan valores bastante estables. Destaca SUPERPOINT, que mantiene un

valor medio de 0,85, el cual es cercano a 1, por lo que está muy bien. En cambio, el que peor resultados tiene es D2NET, ya que presenta un valor medio de 0,4, bastante bajo. El método AKAZE también presenta unos valores bastante estables cercanos al 0,8.

Por otra parte en las secuencias de exterior, figuras 41- 46, observamos valores bajos y grandes picos irregulares. Por ejemplo, si nos fijamos en la figura 36, el método SURF, comienza con valores bajos de 0,2, van incrementando progresivamente hasta llegar a 1 en el pico medio, y a continuación tiene un pico de caída de 0,2 de nuevo, y de repente vuelve a incrementar directamente a 0,6. En este caso, D2NET también presenta valores bajos, entorno al 0,3.

Por lo general, el método que mejores resultados obtiene es SUPERPOINT, ya que además de ser el más estable como podemos observar en la mayoría de las figuras, obtiene los valores más altos lo que demuestra que es más robusto en condiciones de iluminación. Este método mantiene de manera estable los valores entre 0,6 y 0,9.

Por otra parte, el método AKAZE, comienza con valores que van incrementando progresivamente, y después del pico de brillo medio, mantiene valores entre 0,7 y 0,9, como podemos apreciar en las figuras 38,40,41,43,44 y 46.

Los métodos SIFT, BRISK, SURF y ORB, se comportan de manera similar, manteniendo los valores entre 0,4 y 0,6, y con picos en valores antes y después del brillo medio de la imagen, como se puede apreciar en la figura 42.

Finalmente, el método que peores resultados presenta es D2NET, ya que en la mayoría de gráficas, es el que tiene peor rendimiento debido a los picos de caída que tiene y a los valores tan bajos de repetibilidad, que se mantienen entre 0,2 y 0,5, como se puede observar en todas las secuencias. Esto demuestra que tiene una mayor sensibilidad ante cambios de iluminación.

b) *Recall vs Precision*

En este caso, para cada secuencia obtendremos el número de *clusters* obtenidos en cada método, el porcentaje de acierto, los valores de *recall* y los valores de *precision* obtenidos mediante el procedimiento descrito en el apartado 4.2 (ecuaciones 10 y 11),

como se muestra en el ejemplo de la secuencia de interior de la base de datos BASLER en la figura 47. Estos datos serán obtenidos para cada secuencia.

RED	Nº CLUSTERS	PORCENTAJE DE ACIERTO
D2NET	17	1.0
SIFT	47	0.979
ORB	236	0.961
BRISK	66	0.987
AKAZE	91	0.992
SUPERPOINT	307	0.994
SURF	290	0.994

Figura 47. Tabla con el número de clusters y porcentaje de acierto. Para secuencia interior de BASLER

A continuación, se van a mostrar los gráficos obtenidos para las curvas Recall vs Precision con cada descriptor. Para mostrar estos datos, se ha diferenciado entre los resultados obtenidos en las secuencias de la base de datos BASLER y las secuencias ARVC, a su vez, se ha diferenciado entre secuencias interiores y exteriores (Figuras 20 y 21).

Cada gráfica muestra, para cada descriptor, los valores *recall* vs los valores *precision*. Se hace un *zoom* en la parte del final de cada gráfica para poder observar mejor los valores. El objetivo, de nuevo consiste en valorar cómo de estables son los puntos característicos detectados a pesar de los cambios de iluminación en la imagen.

- BASLER
 - Interior

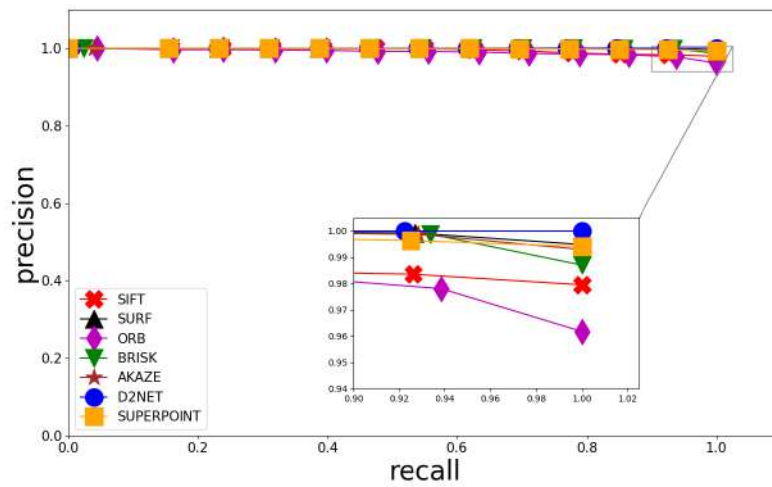


Figura 48. Recall vs Precision, BASLER-database/indoor/sec1

- Exterior

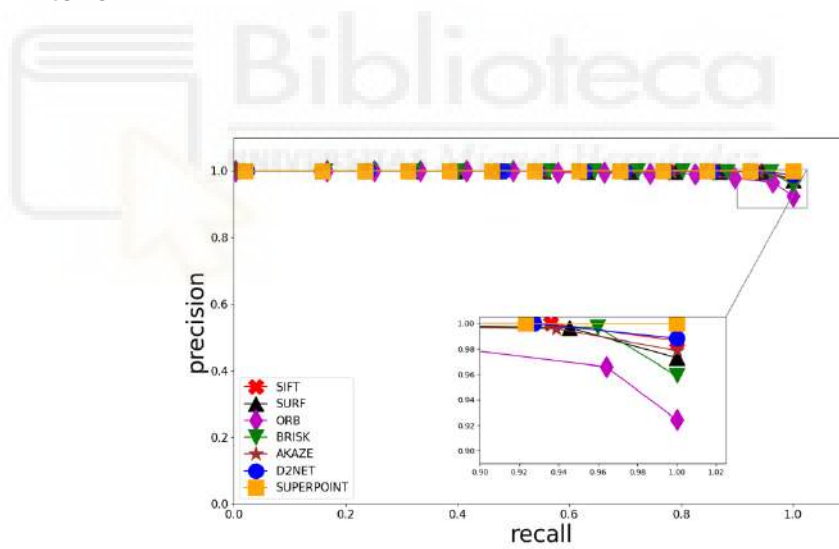


Figura 49. Recall vs Precision, BASLER-database/outdoor/sec1

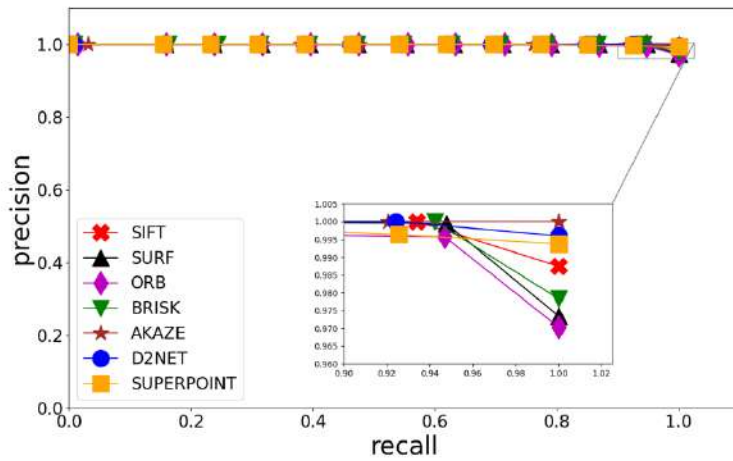


Figura 50. Recall vs Precision, BASLER-database/outdoor/sec2

- ARVC
 - Interior

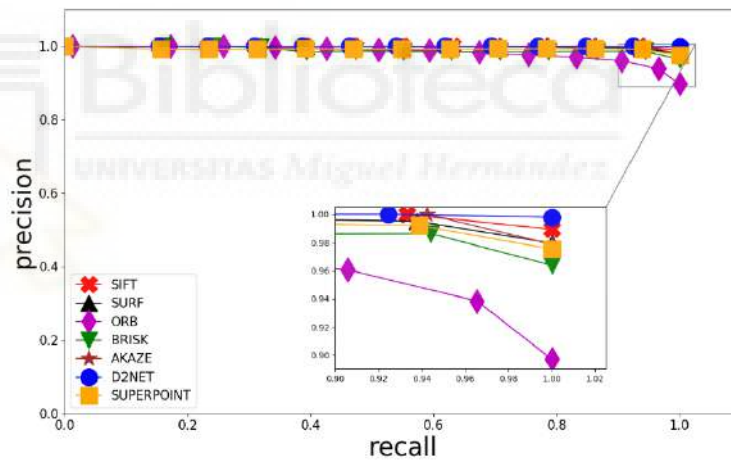


Figura 51. Recall vs Precision, ARVC-database/indoor/sec1

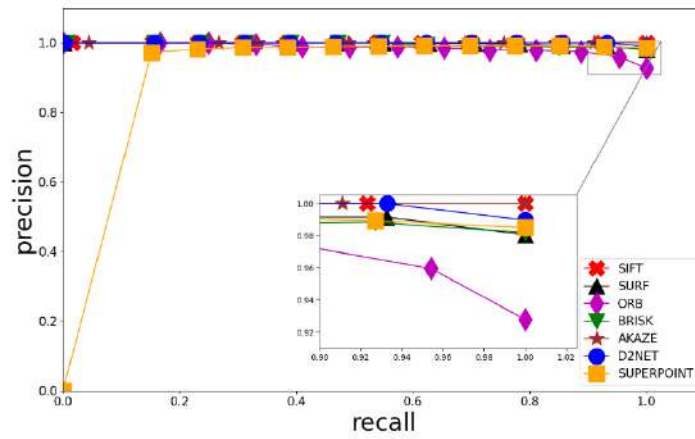


Figura 52. Recall vs Precision, ARVC-database/indoor/sec2

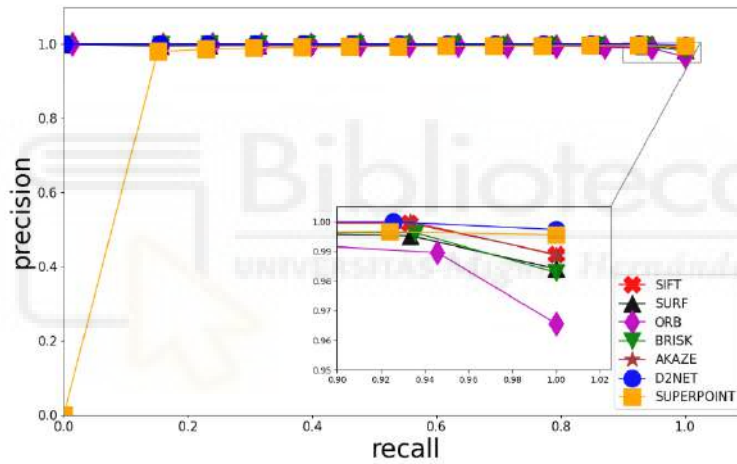


Figura 53. Recall vs Precision, ARVC-database/indoor/sec3

o Exterior

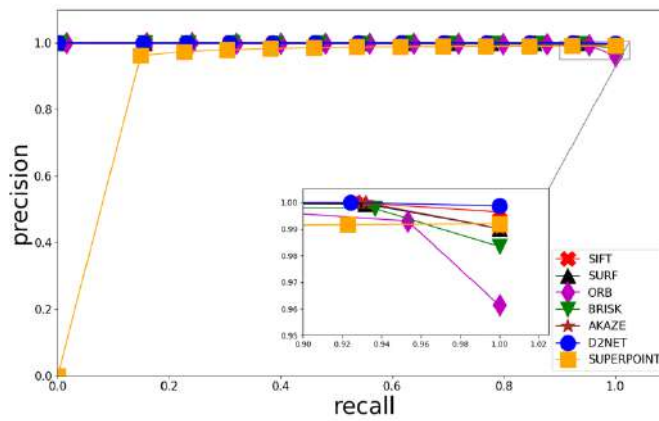


Figura 54. Recall vs Precision, ARVC-database/outdoor/sec1

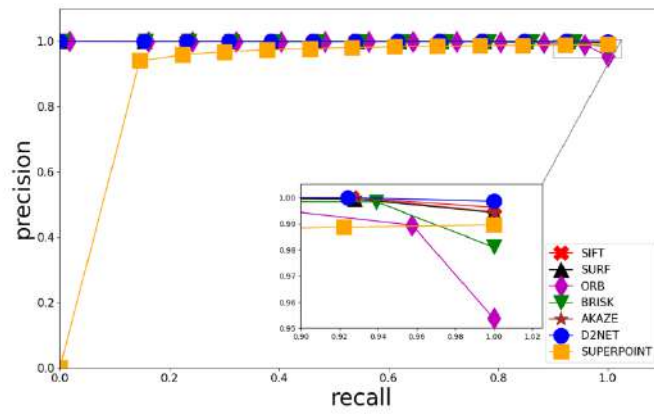


Figura 55. Recall vs Precision, ARVC-database/outdoor/sec2

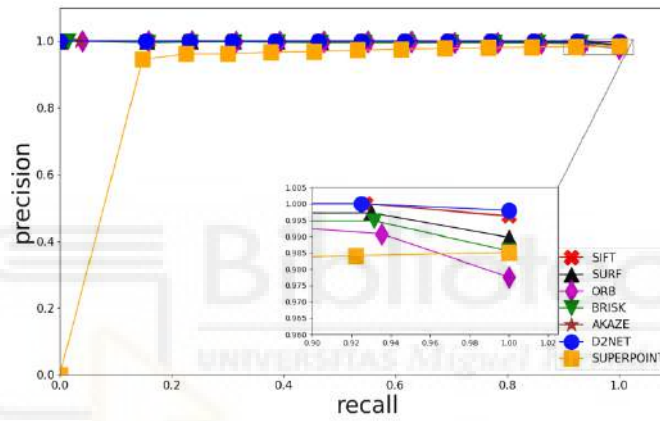


Figura 56. Recall vs Precision, ARVC-database/outdoor/sec3

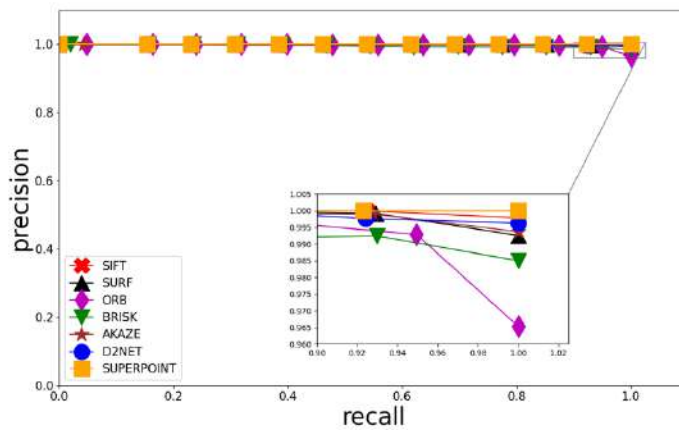


Figura 57. Recall vs Precision, ARVC-database/outdoor/sec4

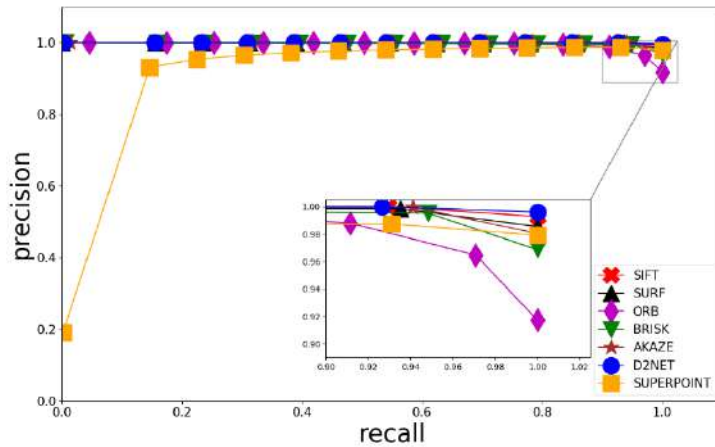


Figura 58. Recall vs Precision, ARVC-database/outdoor/sec6

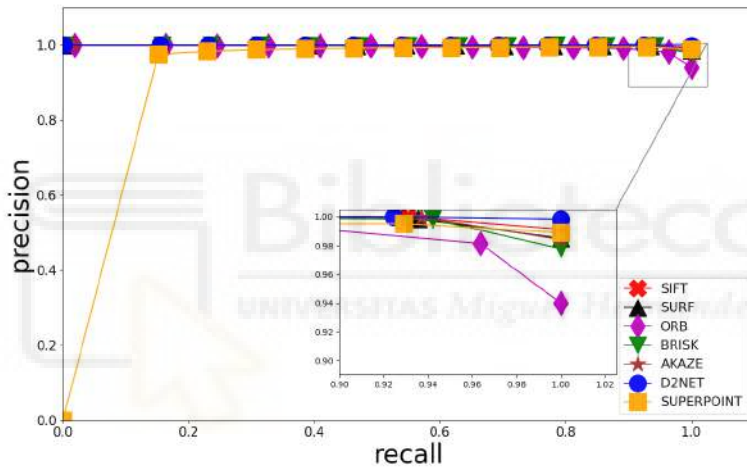


Figura 59. Recall vs Precision, ARVC-database/outdoor/sec7

Como hemos comentado en el apartado 4.2, la situación ideal que buscan las curvas *recall vs precision*, son valores altos tanto de *recall* como de *precision*.

Gracias al *zoom* empleado en las gráficas podemos observar que métodos como SIFT, SURF, BRISK y AKAZE, tienen una ligera caída de los valores de *precision* a medida que aumenta el *recall*. Podemos apreciar esto por ejemplo en la figura 51, a partir del valor 0,94 de *recall*, los valores de *precision* bajan hasta 0,97.

Por otra parte, podemos observar que el método SUPERPOINT, en las figuras 52,53,54,55,56,58 y 59, tarda en alcanzar los valores

cercanos a 1 de *precision*, por lo que es más lento, pero obtiene muy buenos resultados ya que a continuación mantiene los valores en 1.

En cuanto al método que peores resultados presenta, podemos decir que es ORB, ya que en todas las secuencias es el método que tiene mayor caída de los valores de *precision*, descendiendo hasta el 0,89 en ocasiones, como podemos apreciar en la figura 53.

Finalmente, en base a los resultados obtenidos podemos decir que el descriptor que mejor comportamiento tiene es D2NET, ya que en casi todas las secuencias tanto de interior como exterior tiene los valores de *precision* más cercanos a 1 y estables, figuras: 48,51,53,54,55,56,58 y 59.

c) Criterio de Separabilidad

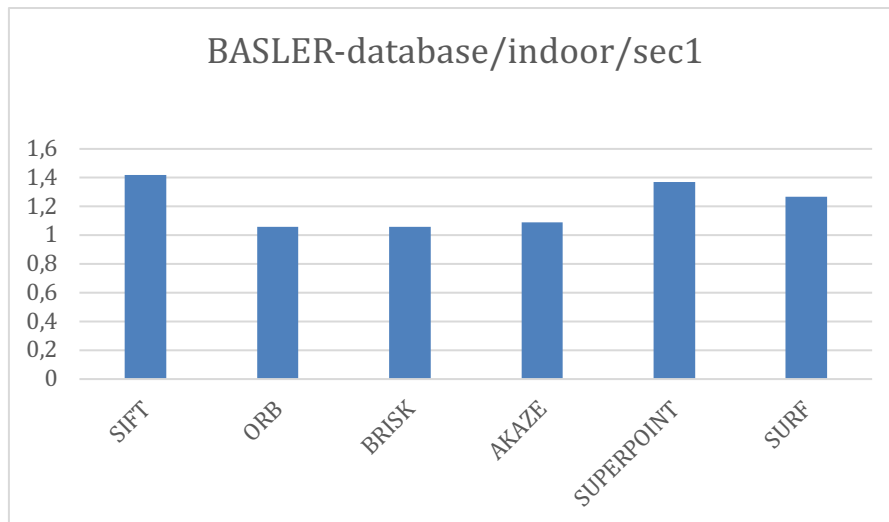
En este caso, evaluaremos los descriptores con el criterio de separabilidad de *clusters*, y para ello calcularemos de la forma descrita en el apartado 4.2, los parámetros $J3$ (ecuación 16), $J3'$ (ecuación 18), Matriz SB (ecuación 14) y Matriz SW (ecuación 12). Además, mostraremos el parámetro $J3$ normalizado en forma de gráfica de barras para poder visualizarlos mejor.

De la misma manera que en las métricas anteriores se ha diferenciado entre los resultados obtenidos en las secuencias de la base de datos BASLER y las secuencias ARVC, y a su vez, se ha diferenciado entre secuencias interiores y exteriores (Figuras 20 y 21).

- BASLER
 - Interior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
D2NET	179086.681	349.778	0.520	2.195
SIFT	181.780	1.420	127089.142	476509.491
ORB	33.834	1.057	132744.062	1110286.059
BRISK	67.707	1.057	360934.632	3777815.900
AKAZE	66.588	1.091	365013.698	2492020.541
SUPERPOINT	351.128	1.371	0.723	1.828
SURF	81.054	1.266	0.362	0.787

a)



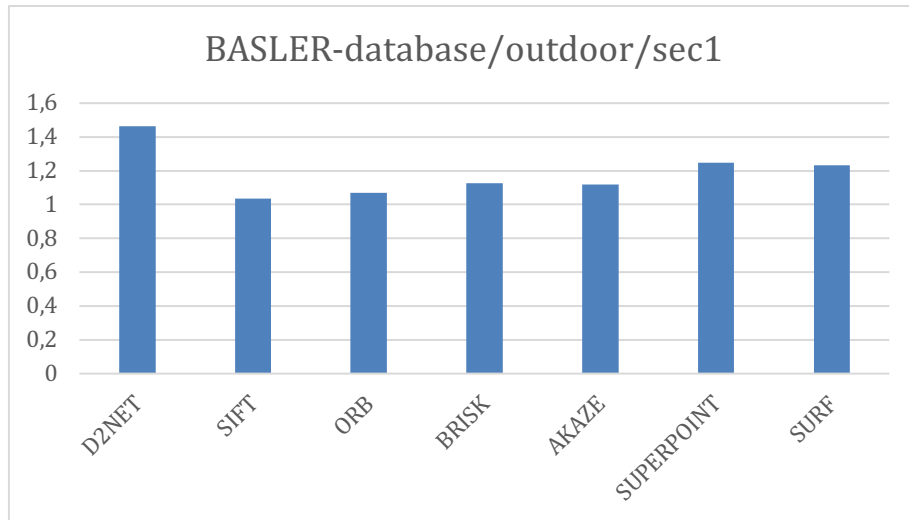
b)

Figura 60. Tabla (a) y gráfica (b) parámetro $J3'$. BASLER-database/indoor/sec1. D2NET no está representado y ha obtenido $J3'=349,778$

o Exterior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER-CLASE	SW MATRIZ INTRA-CLASE
D2NET	750.406	1.465	0.314	2.931
SIFT	132.523	1.035	100801.340	448159.865
ORB	34.196	1.068	105909.169	849207.464
BRISK	72.160	1.127	348261.754	2314605.964
AKAZE	68.201	1.118	311347.392	1932729.096
SUPERPOINT	319.511	1.248	0.674	1.468
SURF	78.793	1.231	0.271	0.514

a)

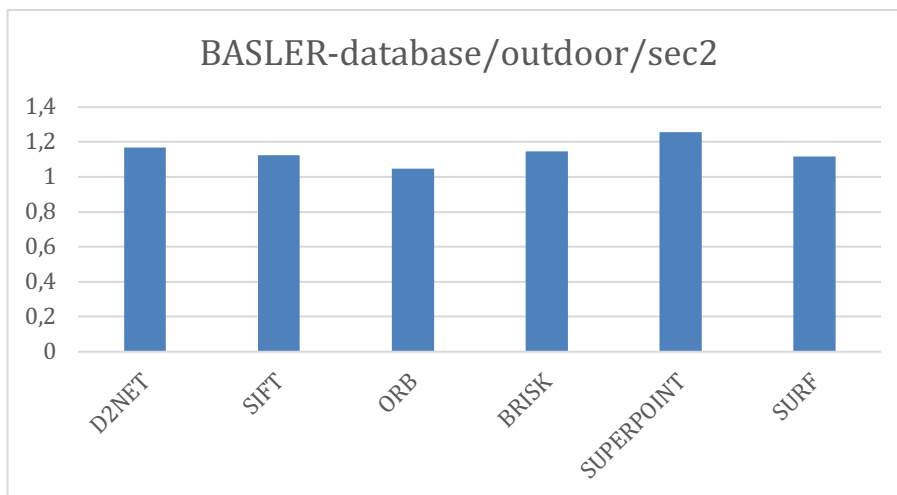


b)

Figura 61. Tabla (a) y gráfica (b) parámetros J3'. BASLER-database/outdoor/sec1

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
D2NET	599,445	1,170	0,405	3,397
SIFT	143,953	1,124	93123,528	948297,047
ORB	33,544	1,048	115537,622	1198223,362
BRISK	73,424	1,147	320404,468	3751176,597
AKAZE	5250,860	86,079	226504,443	2176125,746
SUPERPOINT	322,251	1,258	0,673	2,852
SURF	71,644	1,119	0,245	0,903

a)



b)

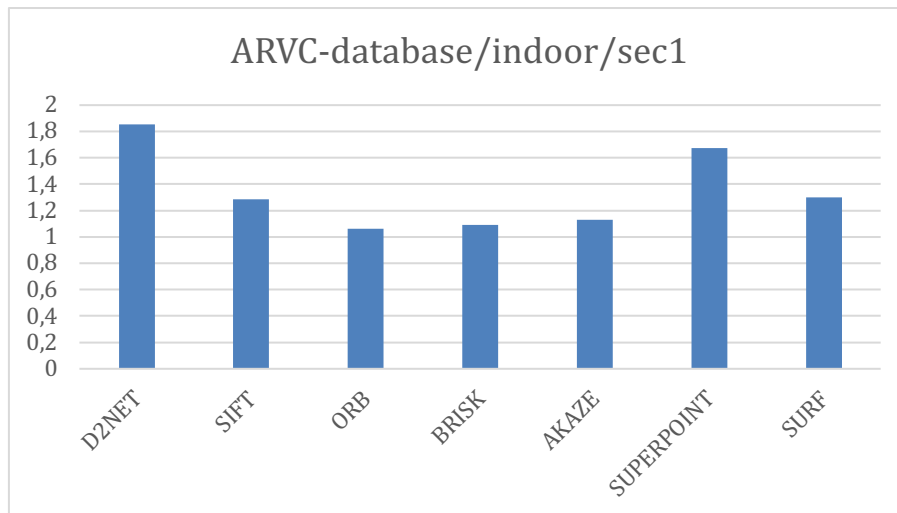
Figura 62. Tabla (a) y gráfica (b) parámetro J3'. BASLER-database/outdoor/sec2. AKAZE no está representado y ha obtenido J3'=86,079

- ARVC

- Interior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER-CLASE	SW MATRIZ INTRA-CLASE
D2NET	949,663	1,854	0,504	0,807
SIFT	164,414	1,284	128394,303	201492,690
ORB	34,063	1,064	106567,231	837133,320
BRISK	69,849	1,091	362953,434	2092440,759
AKAZE	68,931	1,130	307814,633	1816796,250
SUPERPOINT	428,944	1,675	0,713	1,097
SURF	83,197	1,299	0,300	0,600

a)

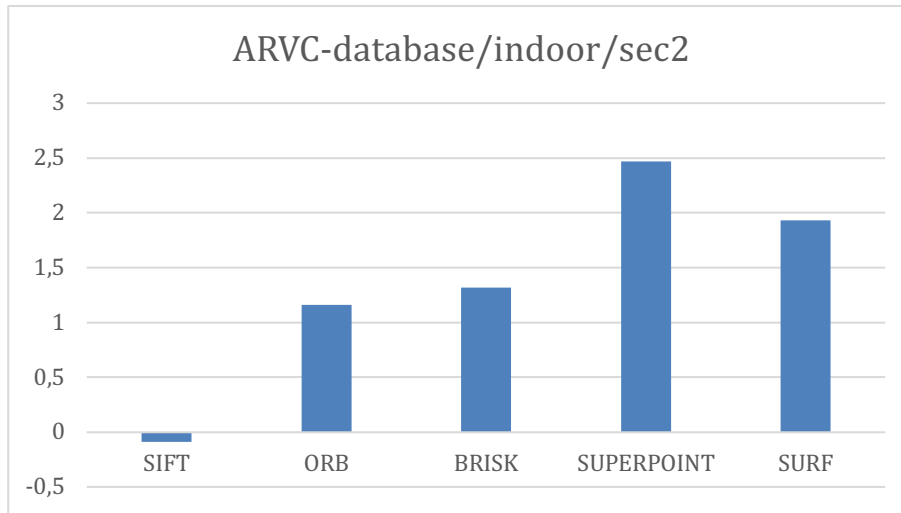


b)

Figura 63. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/indoor/sec1

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
D2NET	-10387,299	-20,287	0,521	0,597
SIFT	-11,389	-0,088	101746,805	196902,273
ORB	37,273	1,164	132781,568	379599,043
BRISK	84,233	1,316	381580,843	1147165,010
AKAZE	31374,096	514,329	344145,262	770680,844
SUPERPOINT	631,955	2,468	0,734	0,659
SURF	123,719	1,933	0,326	0,340

a)

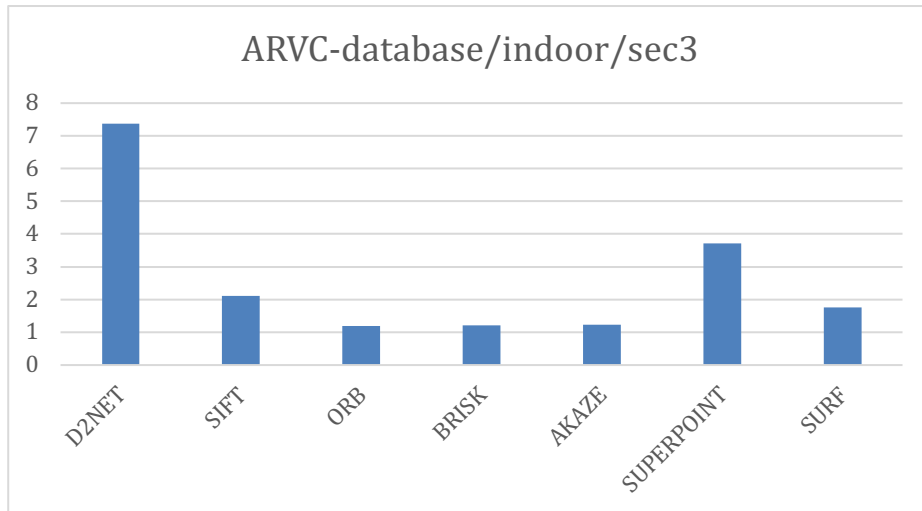


b)

Figura 64.. Tabla (a) y gráfica (b) parámetro $J3'$. ARVC-database/indoor/sec2. AKAZE y D2NET no están representados, y han obtenido: $J3'=514,329$ y $J3'=-20,287$ respectivamente.

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
D2NET	3772,934	7,369	0,563	0,262
SIFT	271,091	2,117	127724,063	172806,759
ORB	38,094	1,190	140927,705	323365,728
BRISK	77,186	1,206	437079,143	872186,340
AKAZE	75,646	1,240	390341,665	697353,091
SUPERPOINT	950,354	3,712	0,810	0,243
SURF	113,043	1,766	0,333	0,249

a)



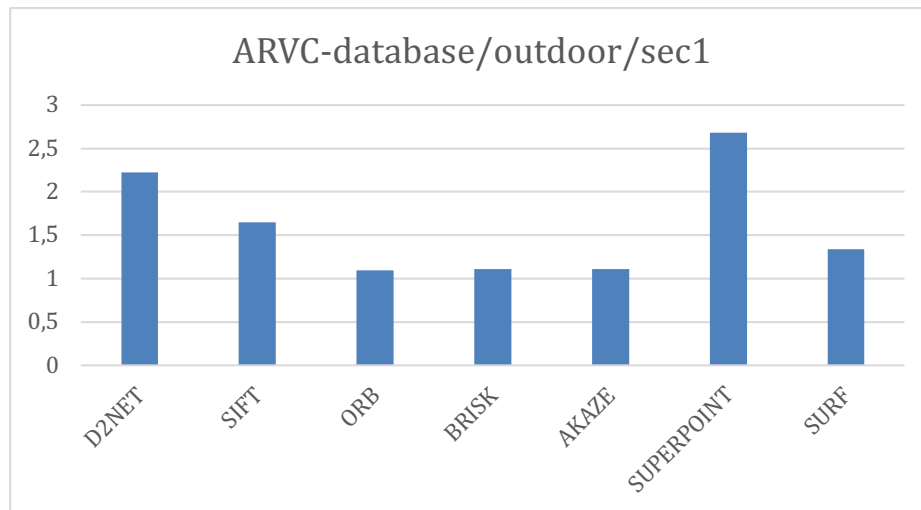
b)

Figura 65. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/indoor/sec3

o Exterior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
D2NET	1137,688	2,222	0,506	0,372
SIFT	210,927	1,647	124202,960	207875,449
ORB	34,993	1,093	131099,185	633930,082
BRISK	70,932	1,108	421235,003	1833206,936
AKAZE	67,683	1,109	383169,962	1497409,465
SUPERPOINT	685,941	2,679	0,653	0,291
SURF	85,902	1,342	0,264	0,387

a)

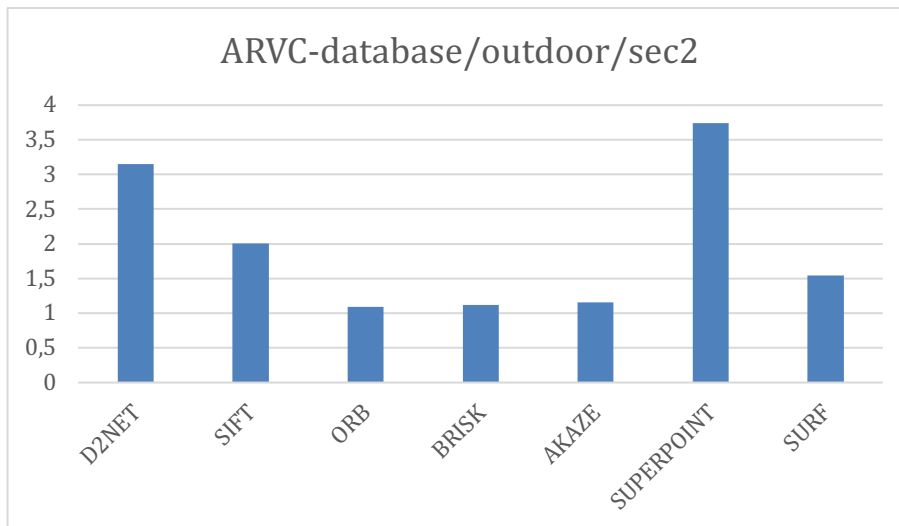


b)

Figura 66. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/outdoor/sec1

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER-CLASE	SW MATRIZ INTRA-CLASE
D2NET	1611,240	3,146	0,541	0,175
SIFT	257,264	2,009	126346,281	145041,873
ORB	35,018	1,094	129489,073	632274,764
BRISK	71,579	1,118	430260,667	1621214,327
AKAZE	70,395	1,154	404931,350	1116289,848
SUPERPOINT	958,457	3,743	0,730	0,171
SURF	99,041	1,547	0,257	0,233

a)

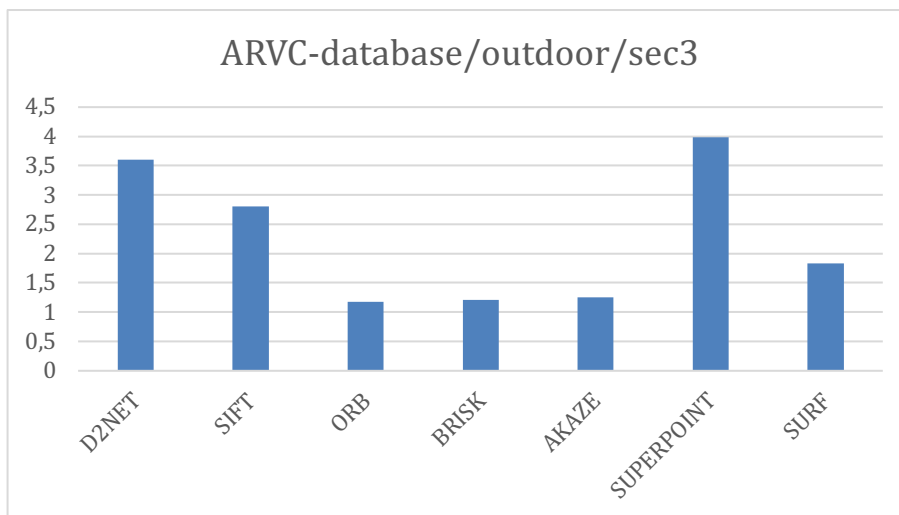


b)

Figura 67. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/outdoor/sec2

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
D2NET	1845,450	3,604	0,599	0,192
SIFT	358,819	2,803	129843,884	169355,570
ORB	37,628	1,175	143225,585	367201,043
BRISK	77,489	1,210	439898,452	982010,407
AKAZE	76,753	1,258	385648,217	767539,006
SUPERPOINT	1019,502	3,982	0,803	0,203
SURF	117,256	1,832	0,364	0,263

a)

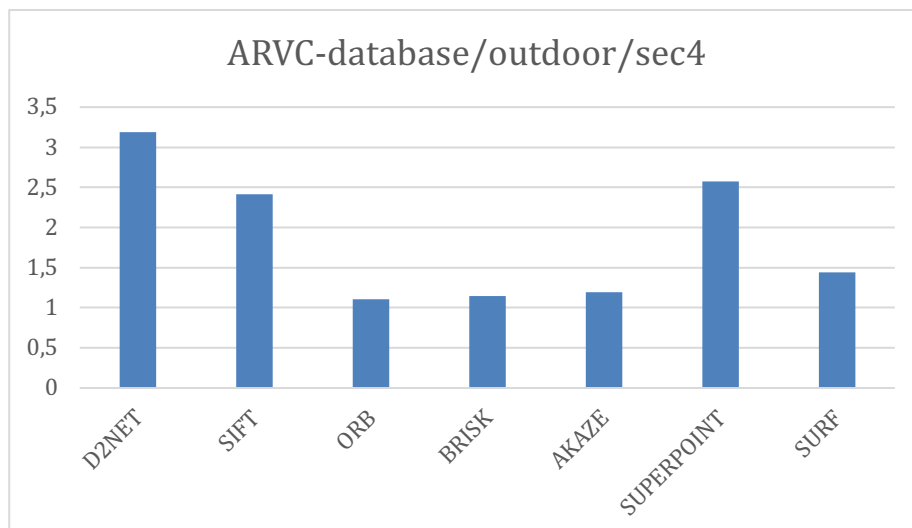


b)

Figura 68. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/outdoor/sec3

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER-CLASE	SW MATRIZ INTRA-CLASE
D2NET	1632,312	3,188	0,546	0,322
SIFT	308,853	2,412	128395,306	168404,065
ORB	35,365	1,105	132599,398	578283,691
BRISK	73,198	1,143	430290,569	1460643,533
AKAZE	72,928	1,195	384818,270	1014347,070
SUPERPOINT	659,073	2,574	0,755	0,381
SURF	92,126	1,439	0,277	0,309

a)

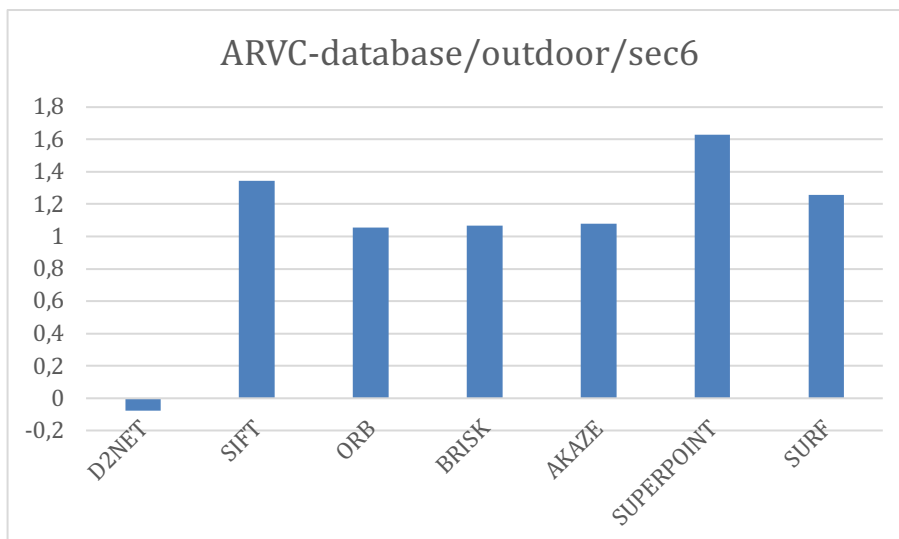


b)

Figura 69. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/outdoor/sec4

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER-CLASE	SW MATRIZ INTRA-CLASE
D2NET	-38,419	-0,075	0,507	0,940
SIFT	172,211	1,345	118143,457	348981,486
ORB	33,740	1,054	112666,780	998349,656
BRISK	68,335	1,067	388159,487	2567854,635
AKAZE	65,814	1,078	359748,952	1947561,774
SUPERPOINT	417,527	1,630	0,694	0,987
SURF	80,481	1,257	0,282	0,503

a)

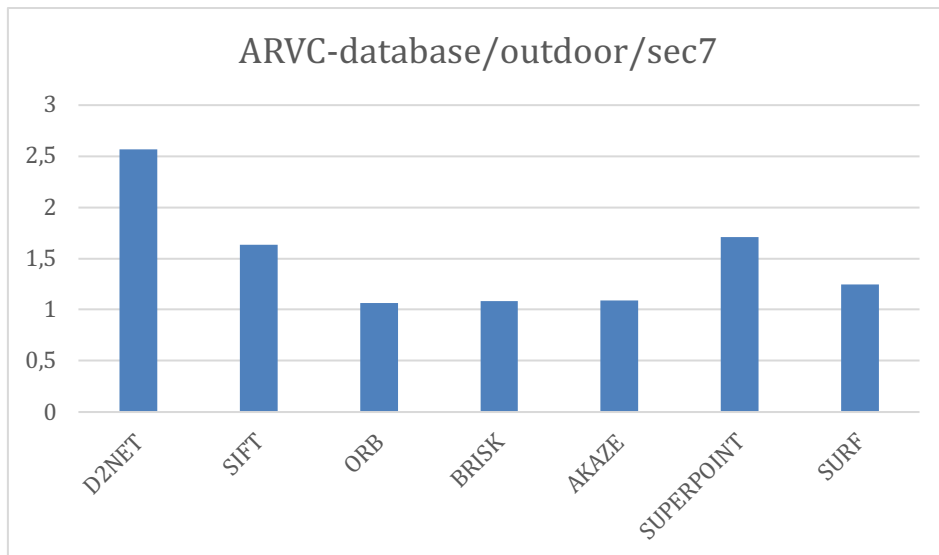


b)

Figura 70. Tabla (a) y gráfica (b) parámetro J3'. ARVC-database/outdoor/sec6

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER-CLASE	SW MATRIZ INTRA-CLASE
D2NET	1314,140	2,566	0,512	0,757
SIFT	209,447	1,636	119032,698	368383,491
ORB	34,021	1,063	117699,040	897606,617
BRISK	69,536	1,086	376834,997	2204418,765
AKAZE	66,438	1,089	343729,658	1830910,463
SUPERPOINT	437,393	1,708	0,695	0,883
SURF	79,622	1,244	0,284	0,537

a)



b)

Figura 71. . Tabla (a) y gráfica (b) parámetro J_3' . ARVC-database/outdoor/sec7

Como hemos comentado en el apartado 4.2, un buen descriptor será aquel que tenga valores bajos de S_w y valores altos de S_b y J_3 .

Únicamente se han representado los métodos que tienen un rango de valores similar, eliminando los métodos que obtienen un J_3' muy elevado, como en las figuras 60, 62 y 64.

En primer lugar, destacan los métodos ORB, BRISK, AKAZE y SIFT, por tener los valores más altos de S_w , por lo que esto nos indica que no presentan buenos resultados en cuanto a separabilidad de *clusters*. Un ejemplo donde se ve claramente es en la figura 63b), gracias al gráfico de barras podemos ver los valores tan altos de la matriz S_w , el método SIFT: 948297, ORB: 1198223, BRISK: 3751176 o AKAZE: 2176125.

Por otra parte los métodos D2NET, SUPERPOINT y SURF, obtienen buena relación de los valores de S_w con S_b y J_3 . Como por ejemplo en la figura 61, podemos observar que el método D2NET, tiene un valor J_3 alto de 179086,681 y un valor S_w bajo de 2,195.

4.3.2. Métodos de características globales

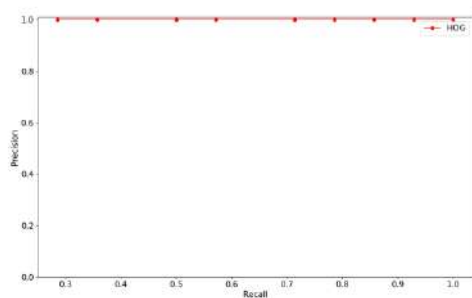
Los métodos que evaluaremos son Firma de Fourier y HOG, descritos en el apartado 3.2.2. En este caso, debido a que por cada imagen de la secuencia habrá un descriptor, un *cluster* estará formado por una secuencia de imágenes.

De la misma manera que en la evaluación de los métodos de características locales, diferenciaremos entre los resultados obtenidos en las secuencias de la base de datos BASLER y las secuencias ARVC, y a su vez, entre secuencias interiores y exteriores (Figuras 20 y 21).

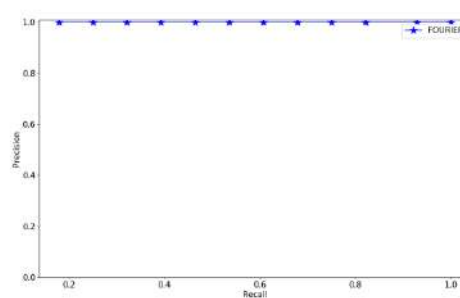
a) *Recall vs Precision*

Igual que con los métodos de características locales, sacaremos los valores *recall* y *precision* para cada método, y a continuación los representaremos en gráficas.

- BASLER
 - Interior



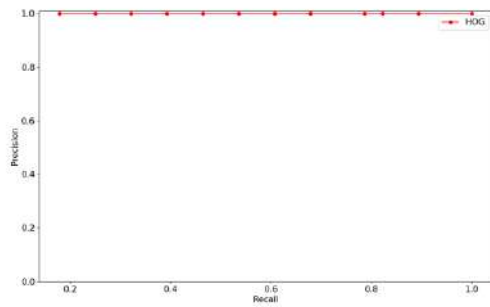
a)



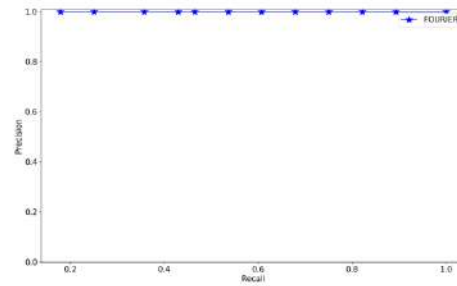
b)

Figura 72. Curva Recall vs Precision BASLER-database/indoor/sec1, 72.a) HOG, 72.b) FOURIER

○ Exterior



a)

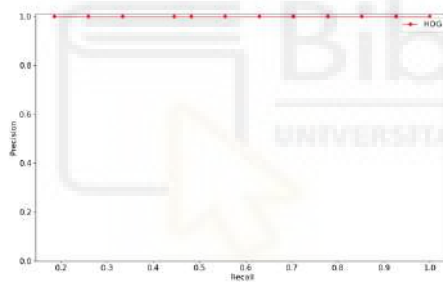


b)

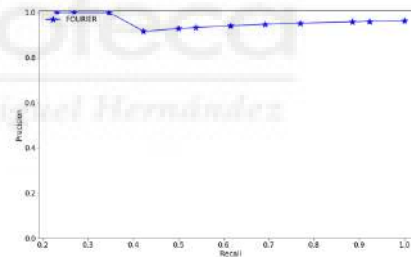
Figura 73. Curva Recall vs Precision BASLER-database/outdoor/sec1-2, 73.a) HOG, 73.b) FOURIER

● ARVC

○ Interior



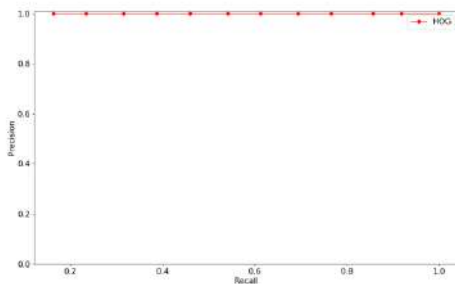
a)



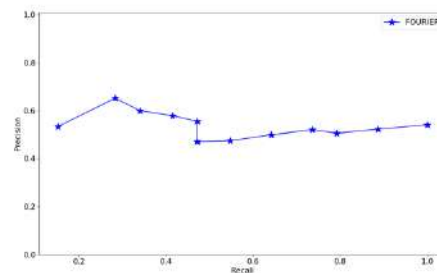
b)

Figura 74. Curva Recall vs Precision ARVC-database/indoor/sec1-3, 74.a) HOG, 74.b) FOURIER

○ Exterior



a)



b)

Figura 75. Curva Recall vs Precision ARVC-database/outdoor/sec1-7, 75.a) HOG, 75.b) FOURIER

Como hemos comentado anteriormente, el descriptor ideal según las curvas *recall vs precision*, son las que alcanzan valores altos tanto en *recall* como *precision*.

Se puede observar fácilmente que el descriptor HOG, obtiene mejores resultados ya que en todas las secuencias alcanza estabilidad en 1 para los valores *precision*, esto se debe a que las correspondencias realizadas basándose en las distancias entre descriptores son siempre correctas.

Por otra parte, en cuanto al método firma de Fourier en la base de datos BASLER (figuras 73 y 74) también alcanza el valor 1, pero al tratarse de pocas secuencias no es un resultado del todo fiable, por lo que en base a los resultados obtenidos en la base de datos ARVC (figuras 75 y 76) podemos observar que este método obtiene resultados más irregulares, sobre todo en las secuencias de exterior, las cuales observamos en la figura 76b) que obtiene valores de *precision* entre 0,4 y 0,6 de manera inestable.

b) Criterio de separabilidad

Una vez más evaluaremos los descriptores con el criterio de separabilidad de *clusters*, pero en este caso para los métodos de características globales, y para ello calcularemos de la forma descrita en el apartado 4.2, los parámetros J_3 (ecuación 16), J_3' (ecuación 18), Matriz SB (ecuación 14) y Matriz SW (ecuación 12).

De la misma manera que en las métricas anteriores se ha diferenciado entre los resultados obtenidos en las secuencias de la base de datos BASLER y las secuencias ARVC, y a su vez, se ha diferenciado entre secuencias interiores y exteriores (Figuras 20 y 21).

- BASLER
 - Interior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA-CLASE
HOG	1,143e+55	1,924e+51	3,255	129,621
FOURIER	49,999	0,999	0,0	2,788

Figura 76. Tabla de parámetros del criterio de separabilidad. BASLER- indoor- sec1

- Exterior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
HOG	1,048e+17	17659240386531,78	3,255	22,366
FOURIER	3,303	0,066	0,090	1,553

Figura 77. Tabla de parámetros del criterio de separabilidad. BASLER- outdoor- sec1-2

- ARVC
 - Interior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
HOG	3,4555e+58	2,165e+54	37,209	32,312
FOURIER	59,250	1,185	0,014	0,075

Figura 78. Tabla de parámetros del criterio de separabilidad. ARVC- indoor- sec1-3

- Exterior

RED	J3	J3 NORMALIZADO	SB MATRIZ INTER- CLASE	SW MATRIZ INTRA- CLASE
HOG	11162,676	0,699	40,265	133,135
FOURIER	51,239	1,024	0,021	0,499

Figura 79. Tabla de parámetros del criterio de separabilidad. ARVC- outdoor- sec 1-7

Como hemos comentado anteriormente, debido a que la base de datos BASLER posee pocas secuencias, nos centraremos más en los resultados obtenidos por la base de datos ARVC.

De la misma manera que para los métodos locales, buscamos obtener valores altos de J3 y Sb, y valores bajos de Sw. En primer lugar, si nos fijamos en la tabla de la figura 78, podemos comprobar que el método HOG, obtiene valores muy altos de J3 y J3' (3,4555e+58 y 2,165e+54), esto quiere decir que es muy sensible a variaciones entre clases. Además, el valor de Sw, 32,212 sugiere que

la dispersión entre clases es considerable, y esto afecta negativamente a la precisión. Finalmente, el valor de S_b , 37,209 indica que las clases están bastante separadas entre sí, lo que es bueno. Por otra parte, en cuanto al método Firma de Fourier, obtiene valores más razonables de J_3 y J_3' (59,250 y 1,18), además de 0,075 de la matriz S_w , lo que implica que tiene una buena precisión.

En siguiente lugar, en las secuencias de exterior, figura 79, el método HOG tiene un valores alto de J_3 como en las de interior, 11162 pero un valor más bajo de J_3' , 0,699. En el caso de la matriz S_w tiene un valor de 133,135 lo que indica una alta dispersión dentro de las clases, y un valor de la matriz S_b de 40,26, que también indica que las clases están bien separadas. Y finalmente el método Firma de Fourier tiene un valor más moderado de J_3 y J_3' , 51, 239 y 1,024 respectivamente, la matriz S_b de valor 0,021 indica poca separación entre clases y S_w 0,499 indica poca dispersión.

Finalmente podemos decir que las secuencias de interior proporcionan mejores resultados para ambos métodos, debido a su mejor capacidad de discriminación (J_3 y J_3') y a su menor dispersión entre clase. Y a su vez, el método Fourier puede ser preferible para situaciones donde la consistencia es preferible a la separabilidad entre clases, y el método HOG, podría ser más adecuado para aplicaciones donde la discriminación entre clases es crucial.

5- CONCLUSIONES.

El objetivo principal de este trabajo de fin de grado fue analizar y comparar diversos métodos de descripción de características utilizados en el campo de la robótica móvil. Esto, se ha realizado con el fin de obtener un buen método para posteriormente poder utilizarlo en el procesamiento de imágenes tomadas por el robot, para crear el mapa por donde se quiera mover.

Para ello, se han evaluado de forma independiente los métodos de descripción de características locales y globales. Los métodos de descripción de características locales que hemos evaluado son SIFT, SURF, ORB, BRISK,

AKAZE, D2NET y SUPERPOINT. Estos dos últimos consisten en redes neuronales. Por otra parte, los métodos de apariencia global evaluados son HOG y Firma de Fourier. En cuanto a la forma de evaluarlos, se han empleado diversos criterios; Criterio de repetibilidad, Curvas *Recall vs precision* y Criterio de separabilidad.

Gracias a los experimentos realizados, hemos podido obtener diferentes resultados. Por una parte, en cuanto a los métodos de descripción de características locales, el criterio ratio de repetibilidad nos ha proporcionado una visión más eficiente del método SUPERPOINT, destacando por ser el más estable y robusto en cuanto a cambios de iluminación. A continuación, el criterio de curvas *recall vs precision*, nos ha proporcionado considerar el método D2NET como el más preciso debido a sus estables resultados. Y finalmente, el criterio de separabilidad de *clusters*, nos ha permitido comprobar que tanto los métodos SUPERPOINT, D2NET y SURF, son los que mejores resultados obtienen en cuanto a la relación de los valores obtenidos. Por lo que por lo general podemos decir que las redes neuronales, SUPERPOINT y D2NET, son las que mejores resultados obtienen debido a la alta precisión y al alto volumen de datos que manejan, a diferencia de los métodos más clásicos como SURF, SIFT, ORB, AKAZE... que obtienen peores resultados para el objetivo que buscábamos en este trabajo.

Por otra parte, en cuanto a los métodos de apariencia global, tanto en el criterio de curvas de *recall vs precision* como el de separabilidad de *Clusters*, el método HOG, obtiene mejores resultados, ya que las correspondencias realizadas basándose en las distancias entre descriptores son siempre correctas. Esto no quita que el método Firma de Fourier también obtiene buenos resultados, pero podría ser utilizado en otro tipo de aplicaciones.

Finalmente, con el objetivo de ampliar este trabajo a futuras investigaciones, se podría centrar únicamente en los métodos basados en las redes neuronales y en encontrar una manera eficiente de entrenarlas utilizando grandes cantidades de datos. Otra investigación interesante sería estudiar la combinación de los

métodos clásicos con los más modernos para aprovechar las cualidades de ambos.

Otra forma de sacarle provecho a este proyecto es probar este proceso completo en un proceso real, es decir, utilizar un robot móvil con una cámara como sensor para que tome fotografías del entorno, aplicar los métodos de detección y descripción, y posteriormente que realice el mapa del entorno y que se localice en él.

Para terminar, una manera de mejorar este trabajo sería ampliando la base de datos BASLER, ya que tiene muy pocas secuencias y por lo tanto los datos obtenidos no son muy fiables.



6. REFERENCIAS

- [1] -Ollero Baturone, Aníbal, "*Robótica manipuladores y robots móviles*", Barcelona Marcombo [2001]
- [2] Blog Robotnik, *Los robots móviles industriales más recientes: usos y aplicaciones*. <https://robotnik.eu/es/los-robots-moviles-industriales-mas-recientes-tendencias-y-usos/>
- [3] Tecnohotel, *Josie Pepper, la robótica asistente personal del aeropuerto de Múnich*.
- [4] Mónica Ballesta, "*Slam visual cooperativo mediante fusión de mapas 3D adquiridos por robots móviles*"
- [5] Universal Robots, "*Sensores en robótica: cuáles usar en sus aplicaciones*"
- [6] Arturo Gil Aparicio, "*Construcción cooperativa de mapas visuales mediante un equipo de robots móviles*"
- [7] John Wang, Edwin Olson, "*AprilTag2: Efficient and robust fiducial detection*"
- [8] Burak Benligiray, "*A stable fiducial marker system*"
- [9] M. Ballesta, A. Gil, O. Reinoso, D. Úbeda, "*Análisis de detectores y descriptores de características visuales en SLAM en entornos interiores y exteriores*"
- [10] Harris, C. G y Stephens, "*A combined corner and edge detector*", 1998
- [11] MIKOLAJCZYK, K. y SCHMID, C.: "*Indexing based on scale invariant interest points*". En: Int. Conf. on Computer Vision, Vancouver, British Columbia, Canada, 2001.
- [12] SMITH, S. M.: "*A new class of corner finder*". En: British Machine Vision Conference, Leeds, Reino Unido, 1992.
- [13] Lowe: "*Distinctive Image Features from Scale-Invariant Keypoints*". International Journal of Computer Vision, 2004, 2(60), pp. 91–110.
- [14] BAY, H.; TUYTELAARS, T. y VAN GOOL, L.: "*SURF: Speeded Up Robust Features*", 2006.

- [15] MATAS, J.; CHUM, O.; URBAN, M. y PAJDLA, T.: “*Robust Wide Baseline Stereo from Maximally Stable Extremal Regions*”. En: In British Machine Vision Conference, pp. 384–393, 2002.
- [16] DAVISON, ANDREW J. y MURRAY, DAVID W.: “*Simultaneous Localisation and Map- Building Using Active Vision*”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002.
- [17] KOSECKA, J.; ZHOU, L.; BARBER, P. y DURIC, Z.: “*Qualitative image based localiza- tion in indoor environments*”. En: Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Madison, WI, USA, 2003.
- [18] ZERNIKE, F.: “*Diffraction theory of the cut procedure and its improved form, the phase contrast method*”. Physica, 1934, 1, pp. 689–704.
- [19] IBM: “¿Qué son las redes neuronales?”, Think 2024.
- [20] Raul Álvarez, “*El director de inteligencia artificial de Tesla nos explica en vídeo como entrenan sus redes neuronales para el uso de Autopilot*”
- [21] Daniel Detone, Tomasz Malisiewicz, Andrew Rabinovich, “*SuperPoint: self-supervised interest point detection and description*”
- [22] Mihai Dusmanu, Ignacio Rocco, Tomas Pajdla y otros : “*D2-NET A trainable CNN for joint Description and Detection of Local Features*”
- [23] Rublee, Ethan. Rabaud, Vincent. Konolig, Kurt. Bradski, Gary “*ORB: an efficient alternative to SIFT or SURF*”, 2013
- [24] Shaharyar Ahmed Khan Tareen; Zahra Saleem, “*Un análisis comparativo de SIFT, SURF, KAZE, AKAZE, ORB y BRISK*”, 2018
- [25] S.Leutenegger et al, “*Binary robust invariant scalable keypoints (BRISK)*”, 2011
- [26] Vicente Román, Luis Payá, Adrián Peidró, David Valiente, Luis Miguel Jiménez, Óscar Reinoso, “*Evaluación de descriptores de apariencia global en tareas de localización bajo cambios de iluminación*”
- [27] L. Payá, L. Fernández, Ó. Reinoso, A. Gil, and D. Úbeda, “*Appearance based dense maps creation-comparison of compression techniques with panoramic images,*” in 2010 IEEE Conference on Emerging Technologies and Factory Automation, ETFA, Bilbao, Spain, September 14-16, 2010 . pp. 250–255.

[28] F. Rossi, A. Ranganathan, F. Dellaert, and E. Menegatti, "Toward topological localization with spherical fourier transform and uncalibrated camera," in Proceedings of the 1st International Conference on Simulation, Modeling and Programming for Autonomous Robots, SIMPAR 2008, Venice, Italy, November 3-6, 2008 . pp. 319–330.

[29] PAYÁ, L., REINOSO, O., BERENGUER, Y., AND U´ BEDA, D. "Using omnidirectional vision to create a model of the environment: A comparative evaluation of global-appearance descriptors". Journal of Sensors 2016 (2016).

[30] Dahi Nemutlu, "Descriptor de características HOG", septiembre, 2022

[31] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, San Diego, CA, USA, June 20-25, 2005 . pp. 886-893.

[32] SIAGIAN, C., AND ITTI, L. "Biologically inspired mobile robot vision localization". IEEE Transactions on Robotics 25, 4 (2009),

[33] RADON, J. 1.1 "Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. Classic papers in modern diagnostic radiology " 5 (2005), 21.

[34] a THEODORIDIS, S. y KOUTROUMBAS, K.: Pattern Recognition. Academic Press, 3

[35] OLIVA, A., AND TORRALBA, "Building the gist of a scene: The role of global image features in recognition." Progress in brain research 155 (2006), 23-36

[36] BERENGUER, Y., PAYÁ, L., PEIDRÓ, A., GIL, A., AND REINOSO, O. "Nerarest position estimation using omnidirectional images and global appearance descriptors" In Robbot 2015: Second Iberian Robotics Conference (2016)

