

# Jornadas de Automática

## Estudio de invariabilidad de marcas visuales ante cambios de iluminación

Máximo, M.<sup>a,\*</sup>, Ballesta, M.<sup>a</sup>, Valiente, D.<sup>a</sup>, Vilella-Cantos, J.<sup>a</sup>, Gil, A.<sup>a</sup>, Payá, L.<sup>a</sup>

<sup>a</sup>Instituto de Investigación en Ingeniería de Elche (I3E), Universidad Miguel Hernández de Elche, Avda. de la Universidad s/n, 03202 Elche (Alicante), España.

**To cite this article:** Máximo, M., Ballesta, M., Valiente, D., Vilella-Cantos, J., Gil, A., Payá, L. 2024. Study of the invariance of visual marks to changes in lighting. *Jornadas de Automática*, 45. <https://doi.org/10.17979/ja-cea.2024.45.10809>

### Resumen

Los robots móviles tienen la capacidad de desplazarse de manera autónoma, para ello es necesario conocer su posición. Esta estimación puede ser obtenida a partir de diversos sensores. Uno de los sensores más comúnmente utilizados son los sistemas de visión por la gran cantidad de información que son capaces de capturar. Para extraer la información más relevante de las imágenes capturadas por las cámaras, se utilizan algoritmos que realizan la detección y su posterior descripción de los puntos más característicos del entorno, también conocidos como marcas visuales. Además, estas marcas visuales deben ser detectadas y descritas de manera que sean invariantes a posibles cambios del entorno. Un caso concreto de estos cambios, son los cambios en la iluminación, común en entornos interiores y exteriores. En este trabajo se ha realizado una evaluación de varios métodos clásicos y métodos que incorporan redes de aprendizaje profundo, para observar su comportamiento ante secuencias de imágenes que presentan diferentes condiciones de iluminación.

*Palabras clave:* robots móviles, percepción y detección, localización, sistemas de posicionamiento, programación y visión

### Study of the invariance of visual marks to changes in lighting

#### Abstract

Mobile robots have the ability to move autonomously, for this purpose it is necessary to know their position. This estimate can be obtained from various sensors. One of the most commonly used sensors are vision systems due to the large amount of information they are capable of capturing. In order to extract the most relevant information from the images captured by the cameras, algorithms are used to detect and describe the most characteristic points of the environment, also known as visual marks. Furthermore, these visual marks must be detected and described in such a way that they are invariant to possible changes in the environment. A specific case of these changes is changes in lighting, which is common in indoor and outdoor environments. In this research, an evaluation of several classical methods and methods incorporating deep learning networks has been carried out to observe their performance when faced with sequences of images presenting different lighting conditions.

*Keywords:* mobile robots, perception and sensing, localization, positioning systems, programming and vision

## 1. Introducción

Una de las características principales que tiene un robot móvil es la capacidad de desplazarse de manera autónoma, para ello uno de los aspectos clave es conocer la posición y orientación del robot. Estos datos son obtenidos a través de sensores embarcados en el robot (Semborski and Idzkowski, 2024).

Dependiendo del sensor seleccionado, la información que se recopila del entorno del robot es diferente. Los sistemas de visión aportan gran cantidad de información acerca del ambiente en el que está situado. En concreto, uno de los sensores de visión más económico son las cámaras. En la bibliografía existe gran cantidad de trabajos que hacen uso de las cámaras como método de localización (Valiente et al., 2017).

La información de las imágenes aportadas por estas cámaras es sintetizada. De esta manera, se extraen las características principales. Estas características pueden ser obtenidas a partir de métodos de detección y descripción.

Existen multitud de métodos clásicos que realizan la extracción de puntos característicos y su posterior descripción, como SIFT (Scale-Invariant Feature Transform) (Lowe, 2004), SURF (Speeded-Up Robust Features) (Bay et al., 2006) y ORB (Oriented FAST and Rotated BRIEF) (Rublee et al., 2011). En la actualidad se han desarrollado diversos algoritmos capaces de detectar y describir puntos de interés en imágenes mediante el uso de técnicas de aprendizaje profundo, como D2-Net (Dusmanu et al., 2019) y SuperPoint (DeTone et al., 2018).

Estos métodos extraen la información visual del entorno en forma de marcas visuales. Una marca visual se corresponde con un punto del espacio. Una de las características que debe poseer una marca visual es que debe ser fácilmente detectable desde diversas posiciones y en múltiples instantes consecutivos. Además, debe ser fácilmente reconocible para el robot, lo que implica que su apariencia visual debe ser descrita de manera que pueda ser distinguida como una misma marca en distintas condiciones. Se han realizado trabajos que estudian la invariabilidad de las marcas obtenidas con métodos clásicos ante cambios en las condiciones del entorno (Gil et al., 2010).

Una de las condiciones del entorno que puede verse modificada es el cambio en la iluminación, pues tanto en los entornos de interior como de exterior se producen variaciones en las condiciones de iluminación en instantes de tiempo diferentes. Por ello, se ha de buscar métodos de detección que sean invariantes a estas modificaciones del entorno y seleccionen las mismas marcas visuales. Además, la descripción de estas marcas debe realizarse de manera que sea independiente a los cambios en la iluminación.

En este trabajo se ha realizado un análisis de distintos métodos de detección y descripción de marcas visuales, y cómo se comportan en imágenes que han sufrido cambios en las condiciones de iluminación.

## 2. Métodos de detección y descripción de características locales en imágenes

Se han analizado diferentes métodos clásicos de extracción de características locales en imágenes tales como SIFT, SURF, ORB, BRISK y AKAZE, además de otros métodos basados en aprendizaje profundo, como son D2-Net y SuperPoint. Todos los métodos seleccionados realizan tanto la detección de puntos de interés en la imagen como su posterior descripción. A continuación se describe el funcionamiento de cada uno de estos métodos.

- SIFT (Lowe, 2004): es un algoritmo en el que los puntos de interés son detectados mediante la búsqueda de máximos locales aplicando la función de diferencia de Gaussianas (DoG) en varias escalas de la imagen de entrada. La descripción es realizada a partir de una ventana de 16x16 píxeles alrededor de los puntos de interés detectados, esta información es sintetizada en un descriptor de tamaño 1x128.

- SURF (Bay et al., 2006): con este método se aplica el determinante de la matriz Hessiana para realizar la detección. Posteriormente, se describe cada uno de los puntos detectados estudiando las respuestas *wavelet* de Haar en las proximidades de los puntos de interés.
- ORB (Rublee et al., 2011): este método combina el algoritmo de extracción de características FAST para la detección con el método BRIEF para la descripción, siendo normalizada en función de la dirección.
- BRISK (Binary Robust Invariant Scalable Keypoints) (Leutenegger et al., 2011): realiza una detección de esquinas mediante el algoritmo AGAST, las cuales son posteriormente filtradas con FAST. La descripción realizada se basa en asignar una dirección a cada uno de los puntos de interés.
- AKAZE (Accelerated-KAZE) (Alcantarilla et al., 2013): este método también se basa en el determinante de la matriz Hessiana para su detección, además de filtros de Scharr para mejorar su invariancia a la rotación. Para la descripción se basa en el algoritmo de Diferencia Local Binaria Modificada.
- D2-Net (Dusmanu et al., 2019): propone la implementación de un método para realizar la detección y la descripción de puntos característicos de manera simultánea. En primer lugar se aplica una red neuronal convolucional profunda (CNN) sobre una imagen de entrada, de manera que se obtiene como salida un tensor 3D. Este tensor se corresponde con un conjunto de mapas de características, a partir del cual se obtienen los descriptores. Estos descriptores son configurados durante el entrenamiento de la red para que un mismo punto del entorno sea representado por descriptores similares. Además, gracias al mapa de características se obtienen los puntos de interés como los puntos cuyo descriptor es localmente diferenciado.

Para inicializar el entrenamiento del método de extracción de características se utiliza la salida de la capa *conv4\_3* de una red neuronal VGG16 preentrenada con el conjunto de imágenes ImageNet. Los datos de entrenamiento se corresponden con imágenes de la base de datos MegaDepth.

- SuperPoint (DeTone et al., 2018): la arquitectura de esta red se corresponde con una red neuronal totalmente convolucional (FCN), y funciona con una imagen completa, en lugar de usando parches de la misma. En este modelo se tiene, inicialmente, un *encoder* que realiza el procesamiento y la reducción de dimensionalidad de la imagen de entrada. Después, se tienen dos *decoder*, uno de ellos entrenado para realizar la detección y el otro para describir los puntos detectados. La red es entrenada previamente con la base de datos MagicPoint, la cuál contiene imágenes sintéticas con formas geométricas.

La implementación de estos métodos se ha realizado utilizando la librería OpenCV para los métodos clásicos. Los repositorios de los métodos D2-Net y SuperPoint se encuentran alojados en GitHub.



Figura 1: imágenes con distintas condiciones de iluminación relativas a las dos secuencias capturadas: (a) Entorno interior; (b) Entorno exterior.

### 3. Base de datos

El estudio de los diferentes métodos de extracción de información de imágenes se ha realizado utilizando datos de varias secuencias de imágenes con una cámara BASLER aCA1600-60GC. Los cambios de iluminación han sido generados mediante cambios en la cantidad de luz que entra a la habitación, y mediante modificaciones en la apertura de la cámara. Las imágenes de una misma secuencia han sido capturadas manteniendo la misma posición y orientación de la cámara, con la finalidad de identificar las correspondencias entre imágenes con cambios de iluminación, basándonos en la posición de los píxeles de la imagen. De esta manera, teniendo secuencias de imágenes estáticas en las que sólo la iluminación se ve modificada podemos evaluar el comportamiento de cada uno de los métodos ante cambios de iluminación, y no ante otras modificaciones como pueden ser cambios en el punto de vista o en la escala.

Se han capturado dos secuencias, una en entornos de interior y otra de exterior, compuestas por 28 y 21 imágenes respectivamente. Estas imágenes son imágenes estándar, y poseen un tamaño de 1600 píxeles de ancho y 1200 píxeles de alto. En la Figura 1 se muestran ejemplos de imágenes pertenecientes a cada una de las secuencias.

### 4. Criterios de evaluación

La evaluación de los métodos de detección y descripción presentados se ha realizado utilizando diversas métricas, con el propósito de valorar el funcionamiento de estos métodos frente a secuencias de imágenes con distintas condiciones de iluminación. Este análisis se ha realizado de manera que permita evaluar de manera independiente tanto la capacidad de detección como de descripción.

El método seleccionado para la evaluación de los detectores de puntos de interés es el denominado **ratio de repetibilidad**. Este parámetro permite evaluar cómo el detector seleccionado funciona extrayendo de manera consistente los mismos puntos del espacio, a pesar de que la escena se vea alterada por variaciones en las condiciones de iluminación.

Para el cálculo de este parámetro, se procede inicialmente a la extracción de características de los puntos de interés empleando cada uno de los métodos seleccionados. Posteriormente, se selecciona una imagen de referencia y se comprueba si un punto detectado en esta imagen aparece en el resto de imágenes de la secuencia. Esta comprobación es posible gracias a que las imágenes han sido tomadas desde una misma posición, lo que implica que la posición en píxeles de un punto en una imagen de la secuencia y en otra deberá ser prácticamente la misma. Una vez hemos obtenido esta información, se define la ratio de repetibilidad según la Ecuación 1, siendo  $np_i$  el número de puntos de interés detectados en la imagen de  $i$  en la secuencia, y  $np_r$  el número de puntos de interés detectados en la imagen de referencia de dicha secuencia.

$$rr_i = \frac{np_i}{np_r} \quad (1)$$

También se ha evaluado la capacidad de descripción de estos métodos mediante la obtención de las curvas de **recall-precision**. Este estudio permite evaluar las asociaciones correctas realizadas entre descriptores. Para esta evaluación se considera que un punto de interés perteneciente a un punto de la escena en las imágenes de la secuencia formará un *cluster* siempre y cuando dicho punto se haya detectado en todas las imágenes de una misma secuencia. De esta manera, un *cluster* estará compuesto por los descriptores de un mismo punto obtenidos para las distintas imágenes de la secuencia. El método de seguimiento de un punto a lo largo de la secuencia se realizará según lo descrito anteriormente, siguiendo los puntos en una misma posición de píxeles en las imágenes. De esta manera, cada uno de los *clusters* que representa el mismo punto de interés  $x$  a lo largo de las imágenes de una secuencia será representado por el conjunto  $D_m^x = \{d_m^x(1), \dots, d_m^x(N)\}$ , siendo  $m$  el método de descripción, y  $N$  el número de imágenes de la secuencia.

A continuación, se obtiene para cada descriptor de un punto su descriptor más cercano entre todos los puntos de todos los *clusters*. La medida de distancia utilizada es la distancia euclídea descrita en la Ecuación 2.

$$E = \sqrt{(d_m^{x_i} - d_m^{x_j})^T (d_m^{x_i} - d_m^{x_j})} \quad (2)$$

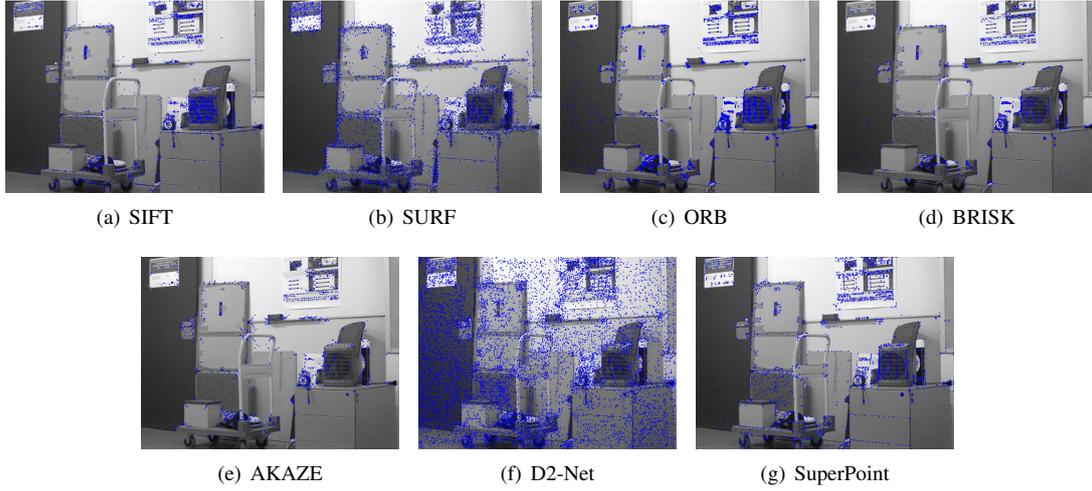


Figura 2: imagen de la secuencia de entornos interiores con los puntos de interés detectados, marcados en color azul, según los diferentes métodos de detección evaluados: (a) SIFT; (b) SURF; (c) ORB; (d) BRISK; (e) AKAZE; (f) D2-Net; (g) SuperPoint.

En este apartado se muestran los resultados obtenidos tras realizar la evaluación de cada uno de los métodos de detección y descripción de características locales en imágenes. Esta evaluación se ha realizado de manera independiente para la detección y para la descripción. Además se ha evaluado el coste computacional de los mismos.

Una vez, obtenidas las distancias mínimas, se comprobará si estas correspondencias son correctas, considerando de esta manera cuando el descriptor  $d_m^{x_i}$  se asigna al *cluster*  $\omega_i$ , y falso cuando se asigna a otro. Con esta información se elabora una lista que incluye cada uno de los descriptores, con una distancia mínima y si está correspondencia es verdadera o falsa. Esta lista se ordenará según los valores de las distancia, y posteriormente se calcularán los valores de *recall* y *precision* según las Ecuaciones 3 y 4 respectivamente estableciendo distintos umbrales de distancias. En estas ecuaciones *num. correspondencias seleccionadas* representa el número de asociaciones en la lista cuya distancia es menor que el umbral establecido, *num. correspondencias correctas seleccionadas* el número de asociaciones correctas del conjunto anterior, y *num. total correspondencias correctas*, el número de asociaciones verdaderas de todo el conjunto.

$$\text{recall} = \frac{\text{num. correspondencias correctas seleccionadas}}{\text{num. total correspondencias correctas}} \quad (3)$$

$$\text{precision} = \frac{\text{num. correspondencias correctas seleccionadas}}{\text{num. correspondencias seleccionadas}} \quad (4)$$

En cuanto a la evaluación del descriptor, también ha de tenerse en cuenta que los descriptores obtenidos además de ser similares para una misma marca visual, deben de ser distintos de los descriptores de marcas diferentes. Para evaluar este factor se ha incluido el **criterio de separabilidad**  $J_3$  (Theodoridis and Koutroumbas, 2006). Para ello, se realiza el cálculo de la matriz intra-clase  $S_w$  y la matriz inter-clase  $S_b$ .  $S_w$  evalúa la invariabilidad del descriptor frente a cambios de iluminación y  $S_b$  evalúa la capacidad de distinción entre clases diferentes.

El cálculo de ambos parámetros se encuentra descrito en las Ecuaciones 5 y 6, donde  $P_i$  la probabilidad de la clase,  $S_i$  es la matriz de covarianza para la clase  $\omega_i$ ,  $\mu_i$  el descriptor medio para la clase  $\omega_i$  y  $\mu_0$  la media global.

$$S_w = \sum_{i=1}^V P_i S_i \quad (5)$$

$$S_b = \sum_{i=1}^V P_i (\mu_i - \mu_0)(\mu_i - \mu_0)^T \quad (6)$$

Una vez obtenidas estas matrices, se obtiene  $J_3$  según la Ecuación 7, donde  $S_m = S_w + S_b$ . Para comparar diferentes descriptores que contienen longitudes diferentes este valor se normaliza según  $J'_3 = \frac{J_3}{L}$ , siendo L la longitud del descriptor.

$$J_3 = \text{trace}(S_w^{-1} S_m) \quad (7)$$

## 5. Resultados

En este apartado se muestran los resultados obtenidos aplicando los criterios de evaluación definidos en el apartado 4, a los métodos de detección y descripción de marcas visuales descritos en el apartado 2. Tal y como se ha comentado en el apartado anterior esta evaluación se ha realizado de manera independiente para la detección y para la descripción de cada uno de los métodos.

### 5.1. Evaluación del detector

En primer lugar se realizó una evaluación de la detección. Tal y como se observa en la Figura 2 cada uno de los métodos funciona de una manera concreta y encuentra los puntos de interés en localización diferentes.

Realizando una evaluación a esta detección se pretende medir la capacidad del descriptor de considerar los puntos de interés en las mismas posiciones a pesar de tener unas condiciones de iluminación diferentes. Para este propósito, se evalúan los puntos de interés de la imagen de referencia que permanecen presentes en el resto de imágenes. Para seleccionar la imagen de referencia, se calcula el brillo medio de cada

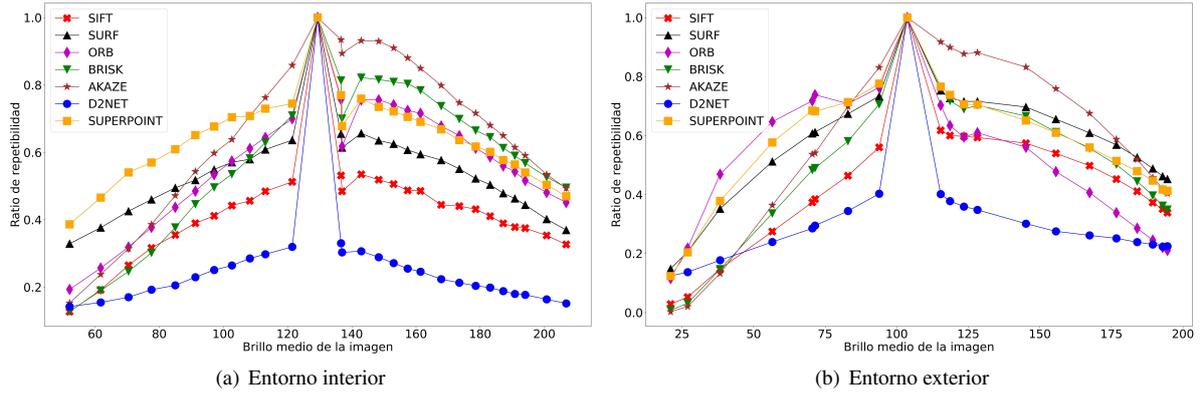


Figura 3: ratio de repetibilidad para las dos secuencias de imágenes con cambios en las condiciones de iluminación utilizadas en este trabajo: (a) Entorno interior; (b) Entorno exterior.

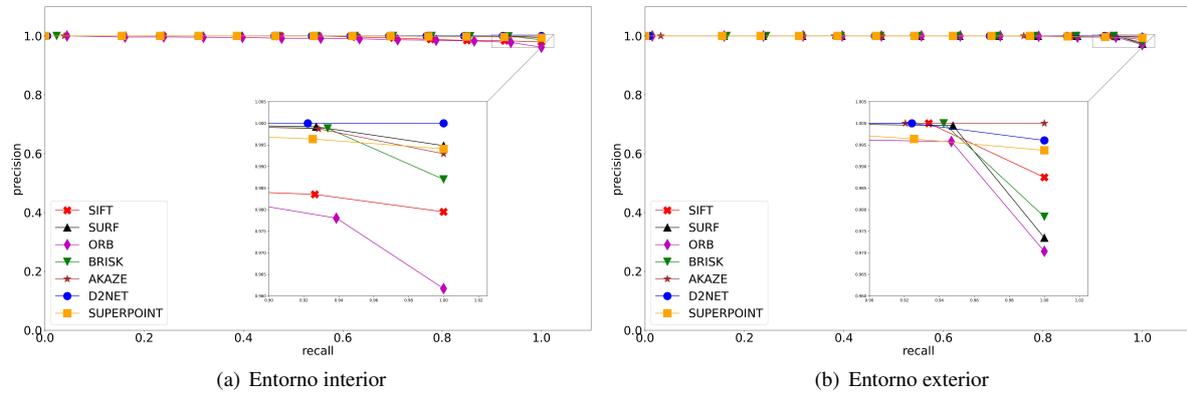


Figura 4: curvas de *recall vs precision* para las dos secuencias de imágenes con cambios en las condiciones de iluminación utilizadas en este trabajo: (a) Entorno interior; (b) Entorno exterior.

una de las imágenes, y se selecciona aquella que tenga un valor intermedio. Después, se obtienen los ratios para cada una de las imágenes. Estos valores se muestran representados en la Figura 3. Esta representación muestra qué descriptores mantienen un valor de la ratio similar en toda la secuencia. En este sentido, D2-Net y SuperPoint son los detectores que ofrecen un mejor comportamiento, ya que para gran parte de las imágenes de la secuencia los puntos detectados son muy similares y se encuentran en las mismas posiciones de la imagen, y se consigue mantener un conjunto de puntos estable en toda la secuencia. Por otra parte, se observa que AKAZE y ORB ofrecen valores de ratio muy diferentes durante la secuencia en ambos casos, lo que significa que cambios en las condiciones de iluminación afectan negativamente a la detección.

### 5.2. Evaluación del descriptor

Por otra parte, con el fin de evaluar la descripción que realiza cada uno de los diferentes algoritmos se han representado las curvas de *recall-precision*. Estas curvas se muestran en la Figura 4. Se observa cómo el valor de precision es mayor a 0.96 para todos los métodos, destacando entre ellos el comportamiento de los métodos D2-Net, SuperPoint y AKAZE. Con esto, se puede concluir que las correspondencias obtenidas a partir de las distancias de los descriptores son en su mayoría correctas para estos métodos.

El algoritmo de descripción también debe cumplir que dos puntos de interés correspondientes a zonas distintas de la ima-

gen deben ser descritos de manera diferente. Esto es, que el *cluster* que representa un mismo punto debe ser compacto y a su vez, estar distanciado de otros *clusters*. Lo que se traduciría en una buena separabilidad. El método de evaluación utilizado para valorar esta propiedad, ha sido el criterio de separabilidad. En la Tabla 1 se recogen los resultados obtenidos de estos parámetros. Los descriptores que mejor comportamiento presentan son los que aportan redes neuronales, es decir, D2-Net y Superpoint, ya que para ambas bases de datos se encuentran entre los métodos con un mayor  $J'_3$ , esto es, realizan la descripción de manera característica para cada *cluster*. En la Tabla 1 también se han incluido los valores de las trazas de  $S_w$  y  $S_b$ , es decir, la suma de las diagonales de estas matrices. Se observa que para todos los métodos los valores relativos a la traza de  $S_b$  son mayores a los de la traza  $S_w$ , por lo tanto, las distancias entre descriptores de *clusters* diferentes son mayores a las distancias entre descriptores de un mismo *cluster*.

### 5.3. Evaluación del coste computacional

Por último, se ha cuantificado el coste computacional de cada uno de los métodos escogidos. Estos valores se recogen en la Tabla 2. En esta Tabla se puede observar como hay métodos computacionalmente más costosos que otros. En concreto, los métodos que involucran redes neuronales requieren una mayor cantidad de tiempo para realizar la detección y descripción, respecto a los métodos clásicos.

	Interior			Exterior		
	$traza(S_b)$	$traza(S_w)$	$J'_3$	$traza(S_b)$	$traza(S_w)$	$J'_3$
<b>SIFT</b>	127089.19	476510.91	1.42	93123.53	948297.05	1.12
<b>SURF</b>	0.36	0.79	1.27	0.25	0.90	1.12
<b>ORB</b>	132744.06	1110286.06	1.06	115537.62	1198223.36	1.05
<b>BRISK</b>	360934.63	3777815.90	1.06	320404.47	3751176.59	1.15
<b>AKAZE</b>	365013.69	2492020.54	1.09	226504.44	2176128.75	86.08
<b>D2-Net</b>	0.52	2.19	349.78	0.41	3.39	1.17
<b>SuperPoint</b>	0.72	1.83	1.37	0.67	2.85	1.26

Tabla 1: criterio de separabilidad obtenido para cada descriptor en cada una de las dos secuencias capturadas.

	t (s)	n° ptos. det.	t(s)/n° ptos.
<b>SIFT</b>	0.1858	2886	0.00006439
<b>SURF</b>	0.1401	5552	0.00002522
<b>ORB</b>	0.0248	7662	0.00000323
<b>BRISK</b>	0.0329	2396	0.00001374
<b>AKAZE</b>	0.0863	1639	0.00005264
<b>D2-Net</b>	35.1068	12984	0.00270384
<b>SuperPoint</b>	2.8915	2738	0.00105607

Tabla 2: coste computacional con CPU para la obtención de puntos de interés y descripción de los mismos en la secuencia de imágenes relativa a entornos de interior. El coste computacional con GPU para los métodos D2-Net y SuperPoint es de 0.058 y 0.149 milisegundos por punto respectivamente.

## 6. Conclusiones

El estudio realizado permite evaluar cada uno de los métodos seleccionados, y observar su comportamiento ante cambios de iluminación en imágenes. Esta evaluación se ha realizado mediante el uso de diversas métricas, las cuales han permitido evaluar de manera independiente la detección de puntos de interés en imágenes y la descripción de los métodos.

Se ha podido observar que para cada caso de estudio, los métodos óptimos son diferentes. Pues en el caso de la detección de puntos de interés, el método SuperPoint encuentra una cantidad de puntos óptima, ya que perviven estos puntos durante toda la secuencia.

En cuanto a la descripción los métodos que mejores resultados ofrecen son D2-Net y SuperPoint, ya que la descripción de los puntos se realiza de manera característica para los puntos de un mismo *cluster*, esto es, los puntos de un mismo *cluster* son descritos de manera similar y a su vez su descriptor es muy distinto del de puntos pertenecientes a otros *clusters*.

Además el estudio se ha realizado con dos secuencias de imágenes capturadas en entornos distintos, tanto de interior como de exterior. Los resultados muestran que los métodos se comportan de manera similar en ambos entornos. Por lo que no existen diferencias significativas en cuánto al comportamiento de cada método en distintos ambientes.

Finalmente, se ha observado que los métodos que aportan redes de aprendizaje profundo presentan un mejor comportamiento, tanto en detección como en descripción. Sin embargo, D2-Net presenta tiempos de computación elevados, ya que detecta un gran número de puntos de interés. Por ello, en cómputo general el mejor algoritmo en términos de detección y descripción ante cambios de iluminación es SuperPoint.

En conclusión, el estudio realizado nos permite observar las diferencias y similitudes entre los métodos evaluados en cuanto a diferentes métricas. De esta manera, se puede selec-

cionar el método que mejor se comporte en cuánto a una propiedad en concreto y a las capacidades de las que se dispone.

## Agradecimientos

Esta publicación forma parte de los proyectos PROMETEO/2021/075, financiado por la Generalitat Valenciana, y TED2021-130901B-I00, financiado por MCI-N/AEI/10.13039/501100011033 y por la Unión Europea “NextGenerationEU”/PRTR

## Referencias

- Alcantarilla, P. F., Nuevo, J., Bartoli, A., 2013. Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces. Proceedings of the British Machine Vision Conference 2013, British Machine Vision Association, Bristol, 13.1–13.11.  
DOI: 10.5244/C.27.13
- Bay, H., Tuytelaars, T., Gool, L., 2006. SURF: Speeded Up Robust Features. Computer Vision—ECCV 2006: 9th European Conference on Computer Vision, Graz, Austria, May 7–13, 2006. Proceedings, Part I 9, 404–417.  
DOI: 10.1007/11744023\_2
- DeTone, D., Malisiewicz, T., Rabinovich, A., 2018. Superpoint: Self-Supervised Interest Point Detection and Description. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition workshops, 224–236.  
DOI: 10.1109/CVPRW.2018.00060
- Dusmanu, M., Rocco, I., Pajdla, T., Pollefeys, M., Sivic, J., Torii, A., Sattler, T., 2019. D2-Net: A Trainable CNN for Joint Description and Detection of Local Features. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition CVPR, 8092–8101.  
DOI: 10.1109/CVPR.2019.00828
- Gil, A., Martínez-Mozos, O., Ballesta, M., Reinoso, O., 2010. A comparative evaluation of interest point detectors and local descriptors for visual SLAM. Machine Vision and Applications, 905–920.  
DOI: 10.1007/s00138-009-0195-x
- Leutenegger, S., Chli, M., Siegwart, R., 2011. BRISK: Binary Robust invariant scalable keypoints. 2011 International conference on computer vision, 2548–2555.  
DOI: 10.1109/ICCV.2011.6126542
- Lowe, D. G., 2004. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, 91–110.  
DOI: 10.1023/B:VISI.0000029664.99615.94
- Rublee, E., Rabaud, V., Konolige, K., Bradski, G., 2011. ORB: An efficient alternative to SIFT or SURF. 2011 International conference on computer vision. IEEE, 2564–2571.  
DOI: 10.1371/journal.pone.0207191
- Semborski, J., Idzkowski, A., 2024. A review on positioning techniques of mobile robots. Robotic Systems and Applications 4, 1–13.  
DOI: 10.21595/rsa.2024.23893
- Theodoridis, S., Koutroumbas, K., 2006. Pattern Recognition, Third Edition. Academic Press, Inc.
- Valiente, D., Reinoso, O., Gil, A., Payá, L., Ballesta, M., 2017. Omnidirectional Localization in vSLAM with Uncertainty Propagation and Bayesian Regression. Advanced Concepts for Intelligent Vision Systems: 18th International Conference, ACIVS 2017, Antwerp, Belgium, September 18–21, 2017, Proceedings 18., 263–274.  
DOI: 10.3390/app7121294