CEA Comité Español de Automática
XLVII JORNADAS DE AUTOMÁTICA CÓRDOBA 2026
UNIVERSIDADE DA CORUÑA

Jornadas de Automática

Localización visual mediante técnicas de vista cruzada con imágenes satelitales

Míriam Máximo[a,*], Judith Vilella-Cantos[a], María Flores[a], Luis Payá[a], David Valiente[a], Mónica Ballesta[a]

[a] *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., Vilella-Cantos, J., Flores, M., Payá, L., Valiente, D., Ballesta, M., 2026. Visual localization using cross-view techniques with satellite imagery. Jornadas de Automática, 47. https://doi.org/10.17979/ja-cea.2026.47.13547

Resumen

El uso de recursos de libre acceso tales como imágenes satelitales utilizadas como mapas para la localización de robots móviles representa una ventaja significativa en términos de eficiencia temporal y económica, ofreciendo una extensa cobertura global sin necesidad de mapeo previo. Para la navegación autónoma de robots móviles se requiere además de una alta precisión en la posición y orientación. Para ello, en este artículo se adapta una arquitectura de aprendizaje profundo originalmente diseñada para estimar la pose bidimensional mediante la correlación de imágenes terrestres y satelitales. En concreto, se ha modificado la información terrestre por proyecciones esféricas generadas a partir de nubes de puntos de un sensor LiDAR 3D. El uso de este sensor aumenta sustancialmente la robustez del sistema frente a variaciones lumínicas, superando las limitaciones de las cámaras convencionales. Se ha realizado una evaluación experimental utilizando la base de datos KITTI, mediante la cual se ha demostrado la viabilidad del enfoque propuesto.

Palabras clave: Percepción y detección sensorial, Procesamiento de imágenes, Programación y visión, Integración de sensores y percepción, Estimación de pose, Aprendizaje profundo, Redes neuronales, Robot móvil

Visual localization using cross-view techniques with satellite imagery.

Abstract

Using open-access resources, such as satellite imagery employed as maps for mobile robot localization, represents a significant advantage in terms of temporal and economic efficiency, offering extensive global coverage without the need for prior mapping. Furthermore, autonomous navigation of mobile robots strictly requires high precision in both position and orientation. To this end, this article adapts a deep learning architecture originally designed to estimate the two-dimensional pose through the correlation of ground-level and satellite images. Specifically, the ground-level input has been modified to spherical projections generated from 3D LiDAR point clouds. The deployment of this sensor substantially enhances the system's robustness against illumination changes, thereby overcoming the limitations of conventional cameras. Finally, an experimental evaluation conducted using the KITTI dataset successfully demonstrates the viability of the proposed approach.

Keywords: Perception and sensing, Image processing, Programming and Vision, Sensor integration and perception, Position estimation, Deep Learning, Neural Networks, Mobile Robot

1. Introducción

La capacidad de mantener una localización precisa y robusta en todo momento constituye uno de los pilares fundamentales de la navegación autónoma. Este problema se aborda mediante técnicas que contrastan las lecturas de sensores embarcados en el robot o vehículo con mapas de alta definición previamente construidos (Macario Barros et al., 2022)(Li et al., 2022). Sin embargo, generar y mantener actualizados estos mapas supone un coste económico y de tiempo muy elevado. Para evitar este problema, el uso de imágenes satelitales

*Autor para correspondencia: mmaximo@umh.es

se presenta como una alternativa eficiente y escalable gracias a su amplia cobertura. Además, plataformas como Google Maps u OpenStreetMap permiten acceder a estas imágenes de forma libre y gratuita.

En la literatura, gran parte de los trabajos basados en imágenes satelitales se enfocan en la localización global del vehículo, proporcionando estimaciones de posición gruesas (Shi et al., 2019) (Hu et al., 2018) (Kang et al., 2025). Dado que la navegación autónoma exige un nivel de precisión estricto en la estimación tanto de la posición como de la orientación, el presente trabajo se centra en el refinamiento de la pose, abordando el problema partiendo de una estimación inicial conocida (Fervers et al., 2022) (Fu et al., 2020).

La percepción del entorno inmediato del vehículo puede llevarse a cabo mediante diversos sensores, presentando cada uno de ellos diferentes desafíos en este contexto. Por ejemplo, las cámaras son altamente susceptibles a las variaciones lumínicas (Alfaro et al., 2025) (Zhang et al., 2021); además, la drástica diferencia de punto de vista respecto a las imágenes cenitales del satélite provoca severas deformaciones al intentar proyectar la información mediante matrices de homografía (Wang et al., 2023). Por el contrario, los sensores LiDAR (Guo et al., 2026) capturan la geometría tridimensional del entorno con alta fidelidad y son invariantes a las condiciones de iluminación, lo que los convierte en una alternativa más robusta para tareas de localización en entornos con condiciones variables. Por ello, en esta investigación se ha optado por el uso de este tipo de sensor, y concretamente nuestro enfoque se basa en la proyección esférica en dos dimensiones de dichas nubes de puntos LiDAR.

Los desafíos que se presentan realizando esta integración entre datos capturados por el robot o vehículo y las imágenes satelitales son los siguientes:

- Variaciones temporales: La presencia de elementos dinámicos o cambios estructurales debido al desfase temporal entre la captura de la imagen satelital y la lectura del sensor a bordo del robot.

- Discrepancia de dominio: El profundo cambio en el punto de vista, ya que la naturaleza de la información contenida en una imagen satelital difiere sustancialmente de la geometría capturada por los sensores terrestres embarcados.

Para dar respuesta a estos retos, este artículo presenta un estudio detallado sobre cómo, a partir de proyecciones bidimensionales de capturas LiDAR, es posible estimar con alta precisión la posición y orientación de un vehículo con respecto a un mapa basado en imágenes satelitales.

Para esta tarea en este trabajo se toma como base la arquitectura Convolutional Cross-View Pose Estimation (CCVPE) (Xia et al., 2023), dónde se formula la estimación de la pose con 3 grados de libertad emparejando los descriptores extraídos de una imagen RGB terrestre con los de un mapa aéreo georreferenciado.

Sin embargo, los enfoques puramente basados en visión son inherentemente vulnerables a variaciones de iluminación y sombras. Para superar estas limitaciones, en nuestra metodología proponemos prescindir de la imagen terrestre y sustituirla por la proyección esférica en dos dimensiones generada a partir de las capturas LiDAR.

De este modo, la codificación de la vista a nivel de suelo no procesa píxeles RGB, sino información correspondiente a la profundidad y la reflectancia del sensor LiDAR proyectadas en la imagen mediante un algoritmo de mapeo. Esta adaptación permite que la estimación de la pose dependa de la morfología del terreno y los obstáculos físicos, mejorando significativamente la robustez de la localización frente a condiciones ambientales adversas, mientras se mantiene la eficiencia computacional del modelo original.

2. Metodología

Inicialmente, tenemos una captura LiDAR del entorno compuesta por puntos p_i con posición (x, y, z) en coordenadas cartesianas, y con un valor de reflectancia r. Además, conocemos una posición aproximada de dónde se encuentra el vehículo, por ello disponemos de una imagen satelital RGB que contiene esta posición. El objetivo es buscar la posición y orientación exactas del vehículo dentro de la imagen satelital. La metodología utilizada se describe en este apartado. Además, este proceso se presenta en la Figura 1.

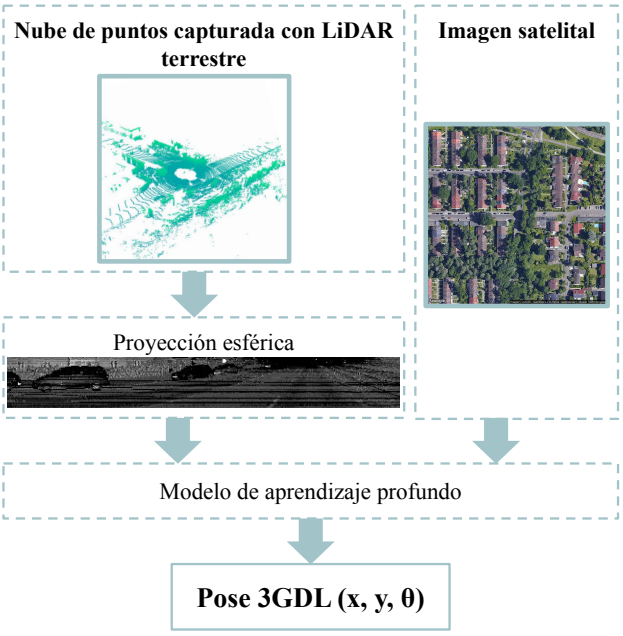


Figura 1: Proceso utilizado para la obtención de la pose. Inicialmente se dispone de una nube de puntos, capturada con el sensor LiDAR terrestre, y de una imagen satelital que contiene la posición dónde ha sido obtenida. A partir de la captura LiDAR se obtiene la proyección esférica en dos dimensiones. Por último, mediante el uso de un modelo de aprendizaje profundo se obtiene la pose del robot.

2.1. *Obtención de la imagen proyectada*

Para procesar la información espacial tridimensional capturada por el sensor LiDAR y permitir su uso en arquitecturas de redes neuronales convolucionales diseñadas para imágenes en 2D, es necesario proyectar la nube de puntos tridimensional en una matriz bidimensional. En este estudio, formulamos

y evaluamos dos enfoques de proyección esférica: el mapeo directo y el mapeo inverso.

El sensor LiDAR utilizado se caracteriza por tener un campo de visión horizontal de 360º y un campo de visión vertical definido por sus límites superior (FOV_{up}) e inferior (FOV_{down}), de modo que el campo de visión vertical total es $FOV = |FOV_{down}| + |FOV_{up}|$, características que se tienen en cuenta para la obtención de la proyección. También se valora el número de haces del LiDAR y la resolución horizontal del LiDAR para fijar el tamaño de la imagen resultante, la cual tendrá una resolución de altura (H) × anchura (W).

Como paso previo común a ambos métodos de mapeo, la nube de puntos original se filtra para eliminar lecturas erróneas u originadas en el propio sensor, conservando únicamente los puntos cuya profundidad d_i se encuentre dentro del intervalo válido $(0, d_{max})$. La profundidad para cada punto se calcula según la Ecuación 1.

$$d_i = \sqrt{x_i^2 + y_i^2 + z_i^2} \tag{1}$$

A continuación, las coordenadas cartesianas se transforman a un sistema de coordenadas esféricas para obtener el ángulo acimutal θ_i y el ángulo de elevación ϕ_i de cada punto p_i según las Ecuaciones 2 y 3.

$$\theta_i = \arctan\left(\frac{y_i}{x_i}\right) \tag{2}$$

$$\phi_i = \arcsin\left(\frac{z_i}{d_i}\right) \tag{3}$$

Posteriormente, se realiza un mapeo directo o inverso dependiendo del caso de estudio, siguiendo el siguiente procedimiento:

- Mapeo directo (MD): En este caso se sigue el procedimiento llevado a cabo en trabajos como (Chen et al., 2020). El enfoque de mapeo directo iterará sobre cada punto p_i de la nube y calculará su posición correspondiente (u_i, v_i) en el plano de la imagen bidimensional. Dado que no se utiliza una proyección equirectangular total que cubra toda la esfera, sino una proyección esférica recortada al campo de visión del sensor, las coordenadas normalizadas continuas se escalan a las dimensiones de la imagen objetivo con las Ecuaciones 4 y 5.

$$u_i = 0{,}5 \cdot \left(1 - \frac{\theta_i}{\pi}\right) \cdot W \tag{4}$$

$$v_i = \left(1 - \frac{\phi_i + |FOV_{down}|}{FOV}\right) \cdot H \tag{5}$$

  Para garantizar que los índices proyectados se mantengan dentro de los límites de la imagen, los valores resultantes se restringen a los intervalos $u_i \in [0, W-1]$ y $v_i \in [0, H-1]$. Un desafío inherente al mapeo directo es la gestión de colisiones, donde múltiples puntos 3D se proyectan sobre el mismo píxel 2D. Para resolver estas oclusiones y preservar la geometría visible desde el origen del sensor, se aplica un ordenamiento descendente de los puntos basado en su profundidad d_i. De este modo, los puntos más distantes se proyectan primero, siendo sobrescritos iterativamente por los puntos más cercanos. Los valores de profundidad y reflectancia se normalizan antes de ser almacenados en la imagen final.

- Mapeo inverso (MI): A diferencia del mapeo directo, que puede generar artefactos de discretización o píxeles vacíos debido a la distribución no uniforme de la nube de puntos, el mapeo inverso garantiza una imagen de proyección densa al iterar sobre la cuadrícula de la imagen destino y buscar el punto más representativo de la nube original. Para cada píxel en la imagen con coordenadas (u, v), donde $u \in \{0, \ldots, W-1\}$ y $v \in \{0, \ldots, H-1\}$, se calculan sus ángulos acimutal $\theta_{u,v}$ y elevación $\phi_{u,v}$ correspondientes según 6 y 7.

$$\theta_{u,v} = \left(1 - \frac{u_i}{W} \cdot 2\right) \cdot \pi \tag{6}$$

$$\phi_{u,v} = \left(1 - \frac{v_i}{H}\right) \cdot FOV - |FOV_{down}| \tag{7}$$

  Una vez obtenidas las coordenadas esféricas de cada uno de los píxeles de la imagen, se buscará el punto más cercano en este sistema entre estas coordenadas $(\theta_{u,v}, \phi_{u,v})$ y (θ_i, ϕ_i). Para ello se emplea una métrica de similitud angular, en concreto se aplicó la métrica de *Haversine* (Sinnott, 1984), mediante la cual se calcula la distancia entre los puntos siguiendo la curvatura de la esfera.

  Para asignar la información a cada píxel (u, v), se selecciona el punto p_i de la nube $\mathcal{P}$ que presente la menor distancia, es decir, la menor desviación angular, con respecto al rayo de proyección proyectado.

  Finalmente, las características del punto seleccionado (profundidad y reflectancia) se asignan al píxel (u, v) para construir la imagen proyectada.

En este trabajo se realiza el estudio de cómo se comporta el algoritmo utilizando exclusivamente la información de profundidad o la de reflectancia. Para ello se asignan estos valores a los tres canales de la imagen utilizando tanto el método de mapeo directo como el inverso descritos. En la Figura 2 puede observarse las imágenes de las cuatro diferentes entradas utilizadas a partir de la misma captura de LiDAR.

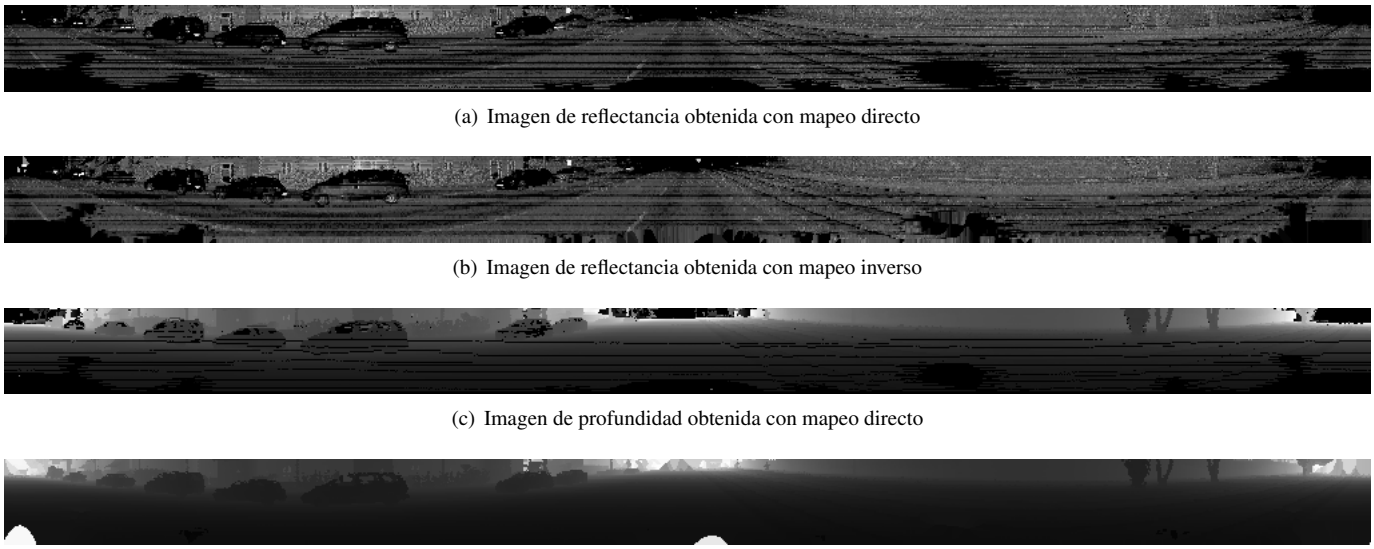
2.2. *Obtención de la pose*

Para la tarea de obtención de la pose, este trabajo adapta la arquitectura Convolutional Cross-View Pose Estimation (CCVPE) (Xia et al., 2023). El objetivo fundamental de esta red es estimar la pose espacial del vehículo con 3 grados de libertad, a partir de la correspondencia entre una vista local a nivel de suelo y un mapa aéreo global georreferenciado.

La arquitectura CCVPE original procesa imágenes RGB terrestres. En nuestra metodología, hemos modificado la entrada para que la red consuma directamente nuestras proyecciones esféricas LiDAR.

Esta arquitectura se inicia con dos ramas de red independientes y sin pesos compartidos que extraen mapas de características multiescala de las entradas a nivel del suelo y aérea. A partir de estos, la red construye un descriptor terrestre unidimensional sensible a la orientación, y una cuadrícula de descriptores aéreos, uno por cada región local del mapa.

El núcleo del sistema es una técnica de rotación y emparejamiento, que desplaza circularmente los descriptores aéreos



(a) Imagen de reflectancia obtenida con mapeo directo

(b) Imagen de reflectancia obtenida con mapeo inverso

(c) Imagen de profundidad obtenida con mapeo directo

(d) Imagen de profundidad obtenida con mapeo inverso

Figura 2: Imágenes obtenidas a partir de la misma captura de LiDAR de la base de datos KITTI, realizando la proyección esférica según el método de mapeo directo e inverso, y utilizando la información de reflectancia y profundidad.

para simular ciertas orientaciones globales posibles y los compara con el descriptor terrestre usando similitud coseno. Este volumen tridimensional de puntuaciones alimenta a dos submódulos especializados. El módulo de localización busca invarianza direccional para saber únicamente su ubicación, lográndolo al extraer de cada píxel exclusivamente la puntuación máxima entre todas las orientaciones posibles. En contraste, el módulo de orientación necesita deducir la dirección el sensor, por lo que omite esta extracción del máximo y preserva el volumen direccional intacto para no destruir la información de los ángulos de rotación.

Finalmente, la información pasa a dos decodificadores separados para generar las predicciones. El decodificador de localización emplea una secuencia de submódulos operando de grueso a fino para producir una distribución de probabilidad espacial densa. Paralelamente, el decodificador de orientación genera un campo vectorial que indica la orientación estimada para cada píxel. La pose definitiva del sistema se infiere seleccionando la coordenada espacial con el valor de probabilidad más alto en el mapa de localización, buscando qué vector direccional se predijo exactamente en ese mismo punto geográfico

La sustitución de la imagen RGB terrestre por la imagen basada en información proveniente de LiDAR es crucial para la robustez del sistema. Mientras que los resultados con la entrada de CCVPE original son vulnerables a variaciones extremas de iluminación, nuestra representación inyecta descriptores puramente estructurales en la red. Esto permite que la correlación cruzada dependa exclusivamente de la topología 3D del entorno, como puede ser la geometría de edificios o la estructura de las carreteras, alineándose de manera mucho más natural con los contornos del mapa aéreo.

3. Experimentos

3.1. Base de datos

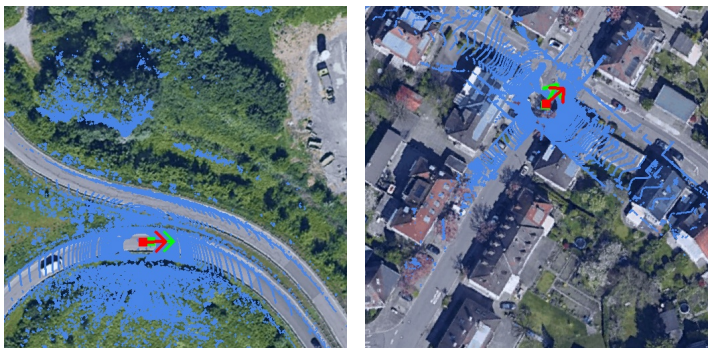
El conjunto de datos empleado en este estudio es KITTI (Geiger et al., 2013). Esta base de datos comprende información recopilada mediante una serie de sensores embarcados en un vehículo, el cual registró diversas trayectorias a lo largo del tiempo. Entre los sensores a bordo destacan un escáner LiDAR 3D Velodyne y un sistema de navegación inercial y GPS de alta precisión. Las secuencias capturadas abarcan una amplia variedad de escenarios, incluyendo entornos urbanos con presencia de edificios, intersecciones y zonas residenciales, así como secuencias capturadas en áreas más periféricas, con mayor presencia de tramos de autovía, vegetación y menor densidad de elementos estructurales urbanos.

Por otro lado, utilizando la información recopilada en (Shi and Li, 2022), se han incorporado imágenes satelitales extraídas de Google Maps, las cuales cuentan con una resolución espacial de aproximadamente 0.2 metros por píxel. Dichos autores dividen los datos de KITTI en tres subconjuntos: uno de entrenamiento (train) y dos de prueba (test1 y test2). El subconjunto test1 contiene datos capturados en regiones geográficamente adyacentes a las zonas de entrenamiento; en contraste, los datos del conjunto test2 pertenecen a áreas más distantes. Esta división tiene como objetivo evaluar la capacidad de generalización del método, permitiendo analizar el rendimiento del sistema tanto en condiciones donde los datos de entrenamiento y prueba comparten similitudes espaciales, como en aquellas donde difieren significativamente.

El LiDAR utilizado en la captura de esta base de datos tiene un campo de visión vertical con límites de 2º y -24.8º, superior e inferior respectivamente, los cuales se tendrán en cuenta para la obtención de las proyecciones de LiDAR. Por otra parte, el tamaño de imagen resultante seleccionado es de altura 64 pixeles, debido a que el LiDAR dispone de 64 haces verticales, por lo que cada fila de pixeles corresponde con un láser. Por otra parte, respecto a la resolución horizontal, el LiDAR posee una resolución de 0.08º. Por lo que en una vuelta de 360º genera 4500 posibles puntos. En este trabajo se ha reducido a una anchura de 1024 pixeles, ya que representa una resolución angular suficiente para no perder detalles estructurales clave, manteniendo un coste computacional bajo para

	test1				test2			
	Error posición (m)		Error orientación (º)		Error posición (m)		Error orientación (º)	
	Media	Mediana	Media	Mediana	Media	Mediana	Media	Mediana
Imagen estándar	6.88	3.47	15.01	6.12	**13.94**	10.98	77.84	63.84
Proy. esférica LiDAR (P, MD)	**1.78**	**0.83**	4.27	2.89	20.11	8.05	66.27	25.88
Proy. esférica LiDAR (R, MD)	2.18	0.88	4.29	2.90	18.24	6.67	61.77	15.21
Proy. esférica LiDAR (P, MI)	2.10	1.05	5.65	4.32	15.47	6.03	63.70	18.66
Proy. esférica LiDAR (R, MI)	2.13	0.88	**3.71**	**2.39**	15.91	**5.89**	**59.90**	**14.82**

Tabla 1: Resultados de los errores de posición y orientación en la base de datos KITTI. Los resultados comparan los resultados del método de referencia CCVPE (Xia et al., 2023) donde una imagen estándar ha sido usada como entrada (mostrados en gris) con los resultados obtenidos usando la proyección esférica de LiDAR haciendo uso de información de profundidad (P) y reflectancia (R) mediante mapeo directo (MD) y mapeo inverso (MI).



(a) (b)

Figura 3: Resultados cualitativos de la localización mediante proyección esférica LiDAR (reflectancia y mapeo inverso). Se muestra la superposición de la nube de puntos proyectada (en azul) sobre la imagen satelital, utilizando la pose estimada para: (a) un escenario del test1 y (b) un escenario del test2. Los cuadrados y las flechas indican la posición y orientación, respectivamente. Se comparan la pose de referencia (en verde) y la pose estimada mediante el proceso de localización (en rojo).

permitir el procesado en tiempo real. Por lo que la imagen resultante es de 64×1024 pixeles.

3.2. Criterios de evaluación

La evaluación cuantitativa del método propuesto se basa en el análisis del error de la pose, desglosado en sus componentes de traslación y rotación. El error de traslación se determina mediante la distancia euclídea entre la posición calculada y la posición verdadera de la base de datos, resumiendo los resultados a través de su media y mediana en metros. Por otra parte, el error de rotación se cuantifica como la diferencia absoluta entre la orientación estimada y la de referencia. Las métricas de error angular resultantes se sintetizarán también mediante sus respectivos valores medios y medianos, expresados en grados.

3.3. Detalles de implementación

El modelo ha sido entrenado durante 10 épocas con un tamaño de lote de 8 y una tasa de aprendizaje de $1{\times}10^{-4}$. En cuanto a la función de pérdida y el optimizador, se han adoptado los mismos que los empleados en la arquitectura CCVPE original (Xia et al., 2023). Todo el proceso de entrenamiento e inferencia se ha llevado a cabo sobre una GPU NVIDIA GeForce RTX 5090.

3.4. Resultados

En este apartado, se exponen y discuten los resultados derivados de los experimentos llevados a cabo para evaluar el desempeño de la metodología propuesta.

Se ha realizado una evaluación aplicando un desalineamiento inicial tanto en la fase de entrenamiento como en la fase de test, siendo este error en orientación dentro del rango ±180 grados, y en posición dentro del rango de ±20 metros tanto en latitud como en longitud. Los resultados obtenidos en los experimentos test1 y test2 utilizando la metodología propuesta, en comparación con el método de referencia (imagen estándar), se presentan en la Tabla 1. Los datos muestran un rendimiento significativamente superior de las proyecciones esféricas LiDAR frente al procesamiento de imágenes estándar en casi todas las métricas evaluadas.

En el primer escenario (test1), las proyecciones basadas en LiDAR logran una reducción drástica del error. Destacando en cuanto al error en posición que el método de mapeo directo con información de profundidad obtuvo los mejores resultados, reduciendo la media del error de posición de 6.88m (imagen estándar) a tan solo 1.78m. En cuanto al error en orientación la configuración de mapeo inverso con información de reflectancia alcanzó el error más bajo, con una media de 3.71º y una mediana de 2.39º, lo que supone una mejora de más del 75 % respecto al método con la imagen estándar (15.01º).

La degradación en el test2 se debe a un gran cambio de dominio respecto al entrenamiento, ya que este conjunto corresponde a áreas más distantes con diferencias estructurales en el entorno. Esta limitación es inherente a los métodos de aprendizaje profundo basados en correspondencia visual. Cabe destacar que, aunque la imagen estándar mantiene una media de error posicional competitiva (13.94m), el mapeo inverso con información de reflectancia demuestra ser el más robusto en términos de estabilidad, logrando la mediana de error de

posición más baja (5.89m) y dominando las métricas de orientación con una media de 59.90º y una mediana de 14.82º.

Por otra parte, se observa que para el test1 el mapeo directo e inverso ofrecen resultados similares. Sin embargo, el mapeo inverso tiende a ofrecer resultados más consistentes y precisos cuando las condiciones se vuelven más exigentes, como en el test2.

En conjunto, el uso de información de reflectancia combinada con el mapeo inverso parece ofrecer el equilibrio más sólido entre precisión y robustez. La metodología propuesta no solo mejora la exactitud métrica, sino que reduce significativamente el error en orientación, validando la eficacia de utilizar proyecciones esféricas LiDAR para tareas de estimación de pose en el dataset KITTI.

La Figura 3 presenta los resultados cualitativos obtenidos mediante el uso de la proyección de reflectancia con mapeo inverso. En el escenario correspondiente al test1, se aprecia una alineación altamente precisa entre la nube de puntos y la imagen satelital, lo que corrobora las métricas de error reducidas reportadas anteriormente. Por el contrario, en el test2 se observa una degradación en la correspondencia espacial; este comportamiento es consistente con el incremento en los errores de posición y orientación analizados anteriormente.

Por último, se ha analizado el coste computacional del sistema. El tiempo de inferencia del modelo es de 13.55 ms de media. En cuanto a la etapa de proyección esférica, el mapeo directo presenta un tiempo medio de 8.88 ms, mientras que el mapeo inverso requiere un coste computacional mayor, con una media de 251.95 ms, debido a su naturaleza iterativa sobre cada píxel de la imagen destino.

4. Conclusiones y trabajos futuros

En este trabajo se ha evaluado una metodología para la estimación de la pose de un robot móvil utilizando imágenes satelitales como mapa para la localización. Para capturar el entorno del vehículo, se han empleado datos de un sensor LiDAR, a partir de los cuales se han generado proyecciones esféricas de los puntos (x, y, z). En este proceso, se han integrado la profundidad y la reflectancia como valores en cada píxel de la imagen resultante.

El método propuesto ha sido validado mediante la base de datos KITTI, la cual integra diversas trayectorias capturadas en entornos y condiciones temporales variadas.

Tras el análisis de los experimentos, se concluye que el uso de la información proveniente del sensor LiDAR mejora los resultados obtenidos con cámaras convencionales. Aunque los resultados obtenidos en el test1 presentan una precisión superior, los resultados en el test2 confirman que se mejora además la capacidad de generalización del modelo ante diferentes escenarios. También, se ha determinado que el uso de imágenes de reflectancia con mapeo inverso ofrece el mejor desempeño global en términos de precisión y robustez.

No obstante, es necesario señalar que, a pesar de las mejoras obtenidas respecto al método de referencia basado en imágenes RGB, los errores de posicionamiento y orientación alcanzados, especialmente en el conjunto test2, no satisfacen aún los requisitos de precisión exigidos por los sistemas de navegación autónoma en entornos reales.

Como trabajos futuros se propone el uso de un método que extraiga las características de las capturas de LiDAR en tres dimensiones, sin realizar proyección previa, aprovechando de esta manera la naturaleza tridimensional del LiDAR.

Agradecimientos

Esta publicación forma parte del proyecto de I+D+i PID2023-149575OB-I00, financiado por MICIU/AEI/10.13039/501100011033 y por FEDER, UE. También ha sido posible gracias al proyecto CIPROM/2024/8, financiado por la Generalitat Valenciana, Conselleria de Educación, Cultura, Universidades y Empleo (programa PROMETEO 2025).

Referencias

Alfaro, M., Cabrera, J. J., Reinoso, Ó., Gil, A., Payá, L., 2025. Localización visual mediante imágenes omnidireccionales y técnicas de fusión temprana. Jornadas de Automática (46).
DOI: 10.17979/ja-cea.2025.46.12239

Chen, X., Läbe, T., Milioto, A., Röhling, T., Vysotska, O., Haag, A., Behley, J., Stachniss, C., 2020. OverlapNet: Loop Closing for LiDAR-based SLAM. In: Proceedings of Robotics: Science and Systems (RSS).
DOI: 10.15607/RSS.2020.XVI.009

Fervers, F., Bullinger, S., Bodensteiner, C., Arens, M., Stiefelhagen, R., 2022. Continuous Self-Localization on Aerial Images Using Visual and Lidar Sensors. In: 2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 7028–7035.
DOI: 10.1109/IROS47612.2022.9982195

Fu, M., Zhu, M., Yang, Y., Song, W., Wang, M., 2020. Lidar-based vehicle localization on the satellite image via a neural network. Robotics and Autonomous Systems 129, 103519.
DOI: https://doi.org/10.1016/j.robot.2020.103519

Geiger, A., Lenz, P., Stiller, C., Urtasun, R., 2013. Vision meets robotics: The KITTI dataset. The international journal of robotics research 32 (11), 1231–1237.
DOI: 10.1177/0278364913491297

Guo, J., Amayri, M., Bouguila, N., Liu, X., Fan, W., 2026. Enhancing rotation-invariant 3d learning with global pose awareness and attention mechanisms. In: Proceedings of the AAAI Conference on Artificial Intelligence. Vol. 40. pp. 4403–4411.

Hu, S., Feng, M., Nguyen, R. M., Lee, G. H., 2018. CVM-Net: Cross-View Matching Network for Image-Based Ground-to-Aerial Geo-Localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7258–7267.
DOI: 10.1109/CVPR.2018.00758

Kang, S., Liao, M. Y., Xia, Y., Wysocki, O., Jutzi, B., Cremers, D., 2025. OPAL: Visibility-aware LiDAR-to-OpenStreetMap Place Recognition via Adaptive Radial Fusion. arXiv preprint arXiv:2504.19258.
DOI: 10.48550/arXiv.2504.19258

Li, L., Kong, X., Zhao, X., Huang, T., Li, W., Wen, F., Zhang, H., Liu, Y., 2022. RINet: Efficient 3D Lidar-Based Place Recognition Using Rotation Invariant Neural Network. IEEE Robotics and Automation Letters 7 (2), 4321–4328.
DOI: 10.1109/LRA.2022.3150499

Macario Barros, A., Michel, M., Moline, Y., Corre, G., Carrel, F., 2022. A Comprehensive Survey of Visual SLAM Algorithms. Robotics 11 (1), 24.
DOI: 10.3390/robotics11010024

Shi, Y., Li, H., 2022. Beyond cross-view image retrieval: Highly accurate vehicle localization using satellite image. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 17010–17020.
DOI: 10.1109/CVPR52688.2022.01650

Shi, Y., Liu, L., Yu, X., Li, H., 2019. Spatial-aware feature aggregation for image based cross-view geo-localization. Advances in Neural Information Processing Systems 32.

Sinnott, R. W., 1984. Virtues of the haversine. Sky and Telescope 68 (2), 159.

Wang, S., Zhang, Y., Perincherry, A., Vora, A., Li, H., 2023. View Consistent Purification for Accurate Cross-View Localization. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8197–8206.
DOI: 10.1109/ICCV51070.2023.00753

Xia, Z., Booij, O., Booij, J. F., 2023. Convolutional Cross-View Pose Estimation. IEEE Transactions on Pattern Analysis and Machine Intelligence 46 (5), 3813–3831.
DOI: 10.1109/TPAMI.2023.3346924

Zhang, X., Wang, L., Su, Y., 2021. Visual place recognition: A survey from deep learning perspective. Pattern Recognition 113, 107760.
DOI: https://doi.org/10.1016/j.patcog.2020.107760