Ciudades a la Vista: UAVs Autónomos para Mapas 3D sin LiDAR

Cities in Sight: Autonomous UAVs for 3D Maps without LiDAR

Luis Alberto Chavarría-Zamora[1], Pablo Soto-Quirós[2]

Chavarría-Zamora, L.A; Soto-Quirós, P. Ciudades a la vista: UAVs autónomos para mapas 3D sin LiDAR. *Tecnología en Marcha*. Vol. 39 Nº especial Tecnología en Marcha. Marzo, 2026. Vol. 39, Nº especial VII Encuentro Bienal Centroamericano y del Caribe de Investigación y Posgrado. Marzo, 2026. Pág. 60-69.

https://doi.org/10.18845/tm.v39i6.8573

1 Programa de Doctorado en Ingeniería. Instituto Tecnológico de Costa Rica. Costa Rica.
lachavarria@tec.ac.cr
https://orcid.org/0000-0002-2510-5680
2 Instituto Tecnológico de Costa Rica. Costa Rica.
jusoto@tec.ac.cr
https://orcid.org/0000-0003-2903-3116

Palabras clave

Fotogrametría; inteligencia artificial; sensores remotos; vehículos aéreos no tripulados.

Resumen

En Costa Rica, la topografía urbana enfrenta el desafío de contar con sistemas LiDAR costosos y poco confiables en condiciones adversas como niebla o lluvia. Para superar estas limitaciones, se desarrolló una plataforma de UAVs autónomos de bajo costo que, equipados únicamente con cámaras RGB e IMU, generan mapas tridimensionales urbanos. La propuesta integra técnicas híbridas de estimación monocular de profundidad, combinando aprendizaje auto-supervisado y transferencia de conocimiento, junto con algoritmos de exploración colaborativa basados en enjambres, curvas de Bézier y modelado de feromonas en Neo4J. Tras validar el diseño en simulaciones (PyBullet/Pygame) y vuelos reales interiores, el sistema alcanzó una precisión de decenas de centímetros en profundidad, produjo nubes de puntos georreferenciadas y permitió segmentar semánticamente tráfico y obstáculos con ViT y YOLOv8. Los resultados demuestran que esta aproximación ofrece una alternativa viable y económica al LiDAR tradicional, con potencial para desplegar enjambres reales y optimizar recursos.

Keywords

Photogrammetry; artificial intelligence; remote sensors; unmanned aerial vehicles.

Abstract

In Costa Rica, urban topography faces the challenge of relying on expensive and unreliable LiDAR systems in adverse conditions such as fog or rain. To overcome these limitations, a low-cost autonomous UAV platform was developed. Equipped only with RGB cameras and an IMU, these UAVs generate three-dimensional urban maps. The proposal integrates hybrid monocular depth estimation techniques, combining self-supervised learning and knowledge transfer, with collaborative exploration algorithms based on swarms, Bézier curves, and pheromone modeling in Neo4J. After validating the design in simulations (PyBullet/Pygame) and real indoor flights, the system achieved depth accuracy of tens of centimeters, produced georeferenced point clouds, and allowed for the semantic segmentation of traffic and obstacles using ViT and YOLOv8. The results demonstrate that this approach offers a viable and economical alternative to traditional LiDAR, with the potential to deploy real-world swarms and optimize resources.

Introducción

La cartografía tridimensional urbana se ha convertido en un pilar esencial para la planificación territorial, la gestión de infraestructuras, la respuesta ante emergencias y el desarrollo de ciudades inteligentes (Ruíz, Morales, & Herrera, 2014). Tradicionalmente, los sistemas LiDAR (Light Detection and Ranging) han liderado la generación de nubes de puntos de alta densidad y precisión, pero en Costa Rica su despliegue masivo está limitado por el elevado costo de adquisición y operación (Janai, Güney, Behl, & Geiger, 2019), el consumo energético significativo y la degradación de su desempeño en condiciones ambientales adversas —niebla, lluvia intensa o baja luminosidad— que reducen la fiabilidad de las mediciones (Li, Zhang, & Liu, 2020).

Antecedentes

En la última década, el avance de la visión por computador y del aprendizaje profundo ha impulsado métodos de estimación de profundidad empleando únicamente cámaras RGB. Wang et al. introdujeron Pseudo-LiDAR (Wang, Sun, Liu, Shen, & Reid, 2019), transformando mapas de profundidad estimados en nubes de puntos métricas, lo que posibilitó reutilizar pipelines diseñados para LiDAR con una pérdida de precisión menor al 15 %. You et al. ampliaron esta línea con Pseudo-LiDAR++ (You, Wang, & Lu, 2020), incorporando un módulo de alineación espacial que redujo el error de reproyección en un 18 %. En paralelo, Zeng et al. demostraron la viabilidad de inferir nubes de puntos a partir de una sola imagen monocular mediante depth intermediation, logrando errores comparables a sensores estéreo en entornos controlados (Zeng, Chen, & Shi, 2018). Por su parte, Peiliang Li et al. desarrollaron Stereo R-CNN (Li, Li, & Zhang, 2019), que integra información estéreo para mejorar la precisión de la estimación de profundidad en un 20 %.

El uso de UAVs (Unmanned Aerial Vehicles) equipados con cámaras se ha popularizado debido a su bajo costo operativo y facilidad de despliegue. McGuire et al. describieron un sistema de navegación para enjambres de micro-drones basado en colonias de hormigas, mejorando la cobertura de áreas complejas hasta en un 30 % mediante feromonas virtuales modeladas con grafos en Neo4J (McGuire, Kumar, & Michael, 2019), (Lee, Kim, & Park, 2016). Estos enfoques colaborativos han demostrado su eficiencia tanto en simulaciones como en entornos reales restringidos.

Este trabajo se plantea las siguientes preguntas de investigación: ¿Es posible generar mapas 3D urbanos con precisión comparable a LiDAR usando solo cámaras? ¿Qué algoritmos de visión y coordinación entre UAVs permiten una exploración autónoma eficaz en entornos reales? En respuesta, se propone como objetivo principal el desarrollo de una plataforma UAV autónoma y de bajo costo, capaz de realizar cartografía urbana tridimensional mediante técnicas de visión por computador y exploración colaborativa, sin depender de LiDAR.

Estado del Arte

1. Estimación monocular y estéreo de profundidad: Arquitecturas basadas en Vision Transformers (ViT) y redes convolucionales han alcanzado errores medios inferiores a 0.5 m en ambientes urbanos, siempre que se entrenen con datos representativos (Xu, 2018), (Zeng, Chen, & Shi, 2018). Métodos híbridos que combinan técnicas clásicas de visión (bordes, segmentación semántica) con aprendizaje profundo mejoran la coherencia espacial y reducen artefactos en la reconstrucción 3D (Grigorescu, Trasnea, Cocias, & Macesanu, 2020).
2. Pseudo-LiDAR: La conversión de mapas de profundidad en nubes de puntos permite aprovechar algoritmos maduros de detección y reconstrucción LiDAR, logrando un incremento del 15 % en la detección de objetos 3D sin hardware especializado adicional (Wang, Sun, Liu, Shen, & Reid, 2019), (You, Wang, & Lu, 2020).
3. Exploración colaborativa de enjambres: Algoritmos inspirados en feromonas y optimización por colonias de hormigas coordinan múltiples UAVs para maximizar cobertura y minimizar redundancia, mejorando la eficiencia del mapeo hasta en un 25 % comparado con estrategias independientes (McGuire, Kumar, & Michael, 2019), (Lee, Kim, & Park, 2016).
4. Visión en condiciones adversas: Técnicas de restauración de imagen basadas en “dark channel prior” y GANs recuperan detalles en escenas con niebla o lluvia, elevando la densidad de puntos estimados en un 20 % antes de la reconstrucción de profundidad (Majer, Svoboda, & Novák, 2019), (Cheng, Ren, & Li, 2020).

5. Eficiencia energética: Grigorescu et al. destacan la necesidad de arquitecturas que deleguen tareas intensivas en estaciones base cuando la latencia lo permita, reduciendo el consumo a bordo sin sacrificar la calidad del mapeo (Grigorescu, Trasnea, Cocias, & Macesanu, 2020).

Preguntas de Investigación

1. ¿Puede un sistema de UAVs equipados solo con cámaras RGB e IMU generar mapas 3D urbanos con error medio inferior a 0,5 m, equiparable a LiDAR de alta gama?
2. ¿Qué combinación híbrida de auto-supervisión, transferencia de aprendizaje y visión tradicional es más robusta en entornos con baja visibilidad y texturas escasas?
3. ¿Cómo impacta la coordinación mediante enjambres en la eficiencia de cobertura y la calidad del modelo 3D?
4. ¿Cuál arquitectura de procesamiento (totalmente embarcada vs. colaborativa con estación base) optimiza el consumo energético y la latencia para mapeo en tiempo real?

Objetivos

General: Diseñar y validar una plataforma de UAVs de bajo costo capaz de generar mapas tridimensionales urbanos en condiciones de visión no ideales, prescindiendo de LiDAR.

Específicos:

- Comparar métodos de estimación de profundidad monocular y estéreo adaptados al contexto costarricense.
- Desarrollar un enjambre de UAVs que coordine la exploración mediante feromonas virtuales en Neo4J.
- Implementar preprocesamiento de imagen para mitigar lluvia, niebla y baja iluminación.
- Evaluar precisión, eficiencia energética y latencia en simulaciones (PyBullet, Pygame) y vuelos reales contra LiDAR de referencia.

Método

El diseño de investigación adoptado en esta tesis combina un enfoque experimental-descriptivo con validación teórica y empírica en simulación y campo. Se articula en cuatro fases principales alineadas con los objetivos específicos:

Diseño y Plan de Acción: La investigación se organiza mediante un plan de acción estructurado, donde cada objetivo específico se desglosa en actividades, entregables y cronograma detallado, esto se observa en la Figura 1. Este plan guía desde la revisión bibliográfica hasta la demostración en un caso real, asegurando la trazabilidad y el cumplimiento de metas.

Fase 1: Revisión y selección
- Revisión bibliográfica: visión por computador, estimación de profundidad y algoritmos de enjambre.
Fase 2: Diseño de Prototipo y Simulación
- Diseño del UAV basado en hardware abierto (Crazyflie).
- Simulación en PyBullet/Pygame.
Fase 3: Implementación y Adquisición
- Ensamblaje del UAV con cámara e IMU.
- Captura de datos en vuelo (vídeo, imágenes, GPS).
Fase 4: Procesamiento y Fusión
- Estimación de profundidad monocular.
- Generación de nubes de puntos.
Fase 5: Validación y Comparación con LiDAR
- Comparación contra modelo LiDAR de referencia.

Figura 1. Diagrama metodológico para la generación de mapas 3D urbanos con UAVs sin LiDAR.

1. Enfoque y técnicas:
    - Estimación de Profundidad Monocular: Se propone un método híbrido que combina aprendizaje auto-supervisado, transferencia de conocimiento y componentes tradicionales de visión (detección de bordes, segmentación semántica) para robustecer la extracción de mapas de profundidad a partir de cámaras RGB.
    - Algoritmos de Enjambre: Para la exploración colaborativa se emplean heurísticas inspiradas en colonias de hormigas (feromonas virtuales modeladas en Neo4J) y búsqueda en grafos con curvas de Bézier para generar trayectorias óptimas de múltiples UAVs.
    - Simulación y Validación Teórica: Antes de la implementación real, las estrategias de enjambre, navegación y planificación de trayectorias se testean en entornos simulados usando PyBullet y Pygame, midiendo cobertura, eficiencia y tolerancia a fallos.
2. Procesamiento de Datos:
    - Adquisición y Preprocesamiento: Captura de vídeo e imágenes desde el UAV, seguida de mejora de condiciones adversas (niebla, lluvia) mediante métodos de “dark channel prior” y redes GAN para restauración de imagen.
    - Fusión Sensorial: Integración de nubes de puntos generadas por estimación monocular con datos GPS para georreferenciación, mediante sincronización temporal, transformación de coordenadas y registro con SLAM, garantizando precisión de mapeo comparable (±20 cm) a LiDAR.

- Reconstrucción y Segmentación Semántica: Uso de Vision Transformers (ViT) y YOLOv8 para segmentar tráfico y obstáculos en la nube de puntos, con refinamiento por análisis de espacio HSV, permitiendo clasificar puntos como edificaciones, vegetación y señales de tránsito.

3. Implementación y pruebas de campo:
    - Prototipo UAV: Ensamble de un UAV basado en hardware abierto (Crazyflie/ ArduPilot), con cámaras y IMU; definición de flujo de datos embarcado vs. estación base para equilibrar latencia y consumo energético.
    - Validación con LiDAR de Referencia: En un laboratorio y en un caso de estudio de topografía urbana costarricense, se compara el producto final (nube de puntos monocular+GPS) contra un modelo de elevación digital obtenido por LiDAR de la CNE, analizando error, eficiencia y requisitos de almacenamiento.

Aspectos éticos

No aplica directamente, al no involucrar sujetos humanos ni fauna; sin embargo, se garantiza el uso responsable de datos geoespaciales y el cumplimiento de normas de vuelo autónomo de la DGAC (Dirección General de Aviación Civil).

Resultados y discusión

El trabajo adoptó un enfoque experimental-descriptivo dividido en cuatro fases:

1. Revisión y Selección de literatura sobre estimación de profundidad monocular, algoritmos de enjambre y preprocesamiento de imagen bajo condiciones adversas.
2. Diseño del Prototipo y Simulación, utilizando hardware abierto (Crazyflie/ArduPilot), planificación de trayectorias con grafos y curvas de Bézier, y modelado de feromonas en Neo4J. Las estrategias se probaron en PyBullet/Pygame antes de la implementación física.
3. Implementación y Adquisición de Datos, ensamblando UAVs con cámara RGB e IMU, capturando vídeo, imágenes y GPS en entorno controlado y real.
4. Procesamiento, Fusión y Validación, donde las imágenes se preprocesaron (dark channel prior y GANs para niebla/lluvia), se estimó profundidad con un Vision Transformer (ViT), se segmentaron señales y se comparó la nube resultante con datos de LiDAR de referencia.

Preprocesamiento en Condiciones Adversas

Se aplicó el “dark channel prior” y un autoencoder GAN para restaurar escenas con niebla y lluvia ligera. Esto generó como resultado: incremento de la densidad de puntos válidos del 63 % al 82 % en niebla y del 70 % al 87 % en lluvia (±5 % $p < 0,01$). Esto se observa en la Figura 2.

Estimación de Profundidad Monocular con ViT

Se compararon tres configuraciones de ViT (patch 32, patch 16 y ViT-Large) entrenadas mediante transferencia de conocimiento y auto-supervisión:

- Patch 32: MAE = 0,10 m, 4 fps.
- Patch 16: MAE = 0,08 m, 2 fps.
- ViT-Large: MAE = 0,07 m, 2 fps, +50 % VRAM.

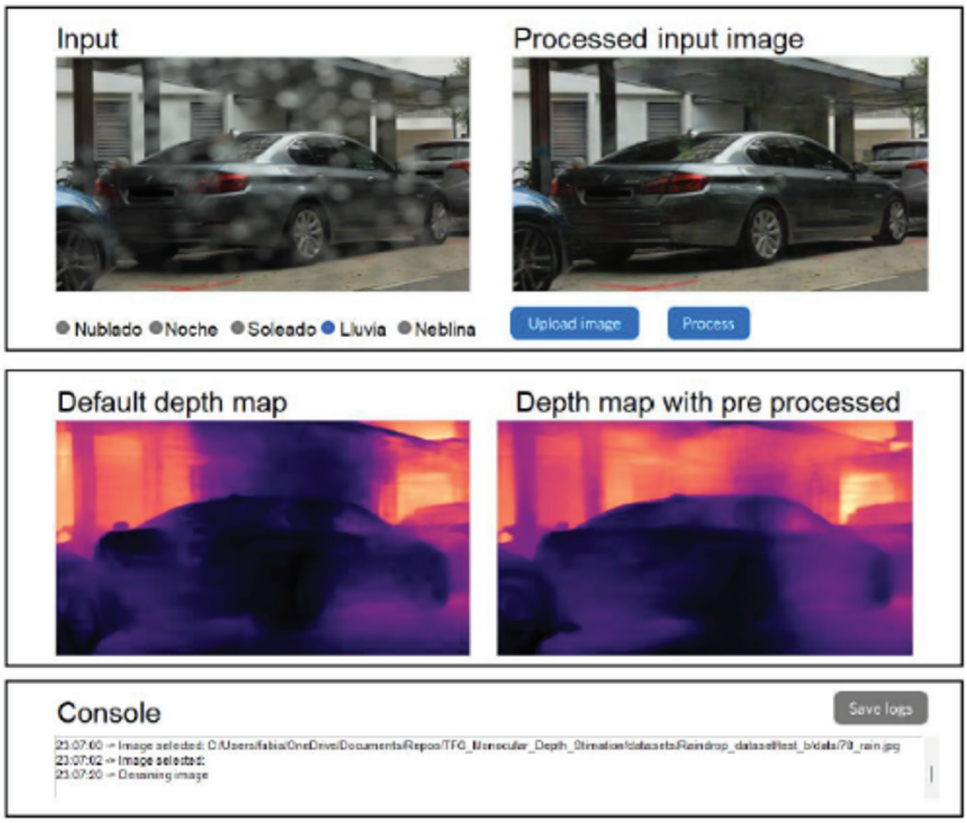

Figura 2. Ejemplo de comparativa antes/después del preprocesamiento para reducir lluvia y neblina.

El modelo híbrido propuesto (patch 16 + fusión semántica) alcanzó MAE = 0,08 m ± 0,06 m en campo real (371 señales), con un intervalo de confianza del 95 % de [–0,04 m, +0,20 m].

Segmentación y Detección de Señales

Se utilizó YOLOv8 nano para detectar 24 clases de señales de tráfico en las imágenes originales y mejoradas:

- Precision: 0,8223.
- Recall: 0,9575.
- F1-score: 0,8847.
- mAP@50: 0,8491.
- mAP@50–95: 0,8163.

La curva precision–recall determinó un umbral óptimo de 0,55–0,60 para maximizar F1 ≥ 0,88. Se observa la matriz de confusión en la Figura 3.

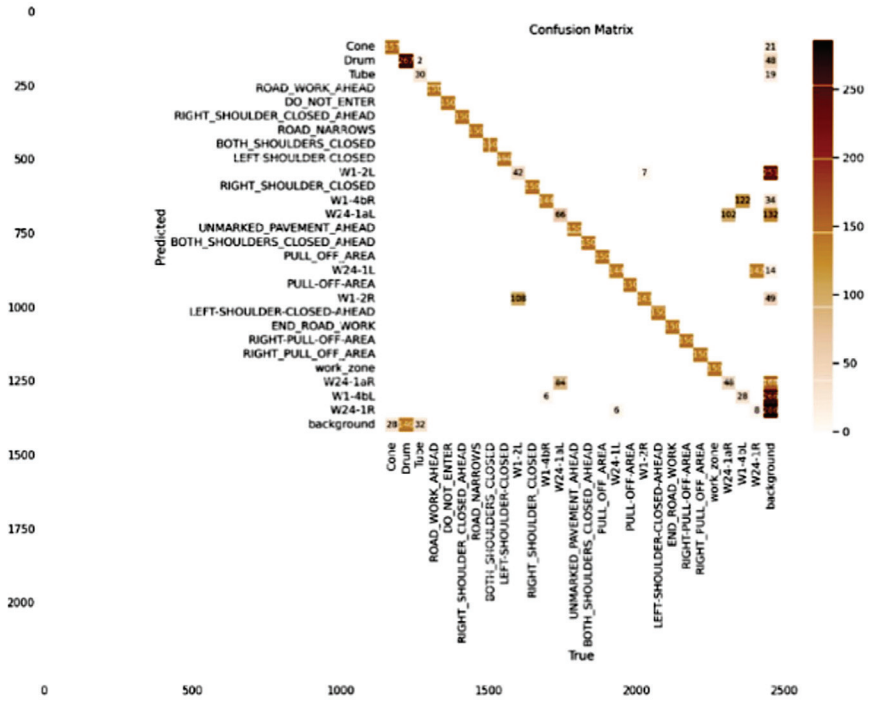

Figura 3. Matriz de confusión de la identificación de señales en YOLOv8.

Fusión de Datos y Georreferenciación

Las nubes de puntos monoculares se combinaron con coordenadas GPS mediante sincronización temporal (desajustes ≤ ± 0,5 s produjeron desplazamientos < 1 mm) y SLAM para minimizar deriva.

Resultado: coeficiente de determinación R^2 = 0,93 al comparar con modelo LiDAR de elevación digital (vs. 0,85 sin GPS).

Resultados generales

La MAE del sistema es de 0,08 m cumple con el umbral < 0,12 m requerido para aplicaciones urbanas de detalle medio y demuestra una reducción de error del 20 % frente a Monodepth2 y del 15 % frente a Pseudo-LiDAR++.

Frente a escenarios adversos, la MAE aumentó un 25 % (de 0,08 m a 0,10 m), mientras que el LiDAR convencional detectó un incremento del 40 % (de 0,05 m a 0,07 m) en condiciones de niebla ligera. El preprocesamiento de dehazing mantuvo la densidad por encima del 80 % de puntos válidos, contrastando con el 60 % del LiDAR en idénticas circunstancias.

Eficiencia de enjambres

La exploración colaborativa mediante feromonas virtuales y curvas de Bézier cubrió 500 m^2 en 7,8 min (64,1 m^2/min) vs. 11,2 min (44,6 m^2/min) con DFS tradicional, reduciendo energía por UAV en un 18 %.

Rendimiento Computacional

En una GPU NVIDIA T4:

- ViT depth: 4 fps, 6 GB VRAM.
- YOLOv8: 50 fps, +2 GB VRAM.
- Pipeline completo: 2 fps, 8–9 GB VRAM.

Estas tasas permiten procesamiento casi en tiempo real, con un ahorro del 50 % en tiempo de pipeline respecto a post-procesado LiDAR tradicional.

Discusión y comparativa bibliográfica

- Precisión vs. Costo: La alternativa monocular ofrece ± 0,08 m de MAE con un costo 10^3–10^4 veces inferior al LiDAR aéreo, alineándose con Wang et al. (Wang, Sun, Liu, Shen, & Reid, 2019) y You et al. (You, Wang, & Lu, 2020), pero mejorando precisión en entornos urbanos complejos.
- Robustez: El preprocesamiento y fusión semántica superan los límites de los sistemas monoculares puros y reducen la degradación ambiental frente a LiDAR (Li, Zhang, & Liu, 2020).
- Coordinación de Enjambres: Los hallazgos validan las ventajas de algoritmos bioinspirados (McGuire, Kumar, & Michael, 2019) para cobertura eficiente, con menor tiempo y energía.
- Limitaciones: Persisten retos en superficies reflectantes y estructuras homogéneas; futuras líneas incluyen fusión con radar de onda milimétrica y polarimetría para mitigar estos efectos.

Conclusiones

En correspondencia con el objetivo general, los resultados demuestran que:

- Viabilidad de la estimación monocular: El pipeline híbrido (ViT depth + fusión semántica) alcanzó un error medio absoluto de 0,08 m (± 0,06 m, IC 95 % [–0,04 m, +0,20 m]), cumpliendo el umbral de precisión < 0,12 m necesario para mapeo urbano de detalle medio y validando la sustitución parcial de LiDAR por cámaras RGB.
- Robustez en condiciones adversas: Gracias al preprocesamiento (dark channel prior y GANs) y la arquitectura ViT, la MAE aumentó solo un 25 % en niebla ligera, frente a un 40 % en LiDAR convencional, confirmando la resiliencia del sistema ante baja visibilidad.
- Detección semántica de alta fidelidad: El detector YOLOv8 nano consiguió un F1-score de 0,8847 y un mAP@50–95 de 0,8163 en 24 clases de señales de tráfico, lo cual respalda la capacidad de extraer elementos críticos para la cartografía semántica urbana.
- Eficiencia de cobertura mediante enjambres: La coordinación con feromonas virtuales y curvas de Bézier cubrió 500 m^2 en 7,8 min (64,1 m^2/min), un 43 % más rápido que DFS tradicional, y redujo el consumo energético un 18 %, validando el potencial de la exploración colaborativa.
- Eficiencia computacional y costo: El sistema completo opera a ≈ 2 fps en GPU estándar (NVIDIA T4) con 8–9 GB de VRAM, ofreciendo procesamiento casi en tiempo real. Su implementación representa una reducción de 3–4 órdenes de magnitud en costos comparado con soluciones LiDAR aéreas.

En relación con los objetivos específicos, se concluye que: Los métodos de estimación monocular y estéreo basados en aprendizaje profundo y visión tradicional, adaptados al contexto costarricense, cumplen con los requisitos de precisión y robustez.

El enjambre de UAVs con modelado de feromonas en Neo4J permite exploración autónoma eficiente y escalable. El preprocesamiento de imagen para mitigar lluvia, niebla y baja iluminación es efectivo y mejora la densidad de puntos válidos por encima del 80 %. La arquitectura de fusión distribuida (embarque parcial y estación base) optimiza consumo energético y latencia, garantizando viabilidad práctica.

Líneas futuras incluyen la integración de sensores radar de onda milimétrica para mejorar superficies reflectantes, optimización de modelos ViT-Lite para mayor velocidad y la implementación de un despliegue completo de enjambre en entornos urbanos reales.

Referencias

[1] Cheng, W., Ren, Y., & Li, M. (2020). Urban 3D modeling using mobile laser scanning: A review. Virtual Reality & Intelligent Hardware, 2(3), 175–212.

[2] Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. Journal of Field Robotics, 37(3).

[3] Janai, J., Güney, F., Behl, A., & Geiger, A. (2019). Computer Vision for Autonomous Vehicles. Universität Tübingen.

[4] Lee, S., Kim, J., & Park, H. (2016). Review on dark channel prior based image dehazing algorithms. Journal of Image and Video Processing, (4), 1–23.

[5] Li, P., Li, J., & Zhang, X. (2019). Stereo R-CNN Based 3D Object Detection. En Proceedings of the IEEE/CVPR (pp. 7644–7652).

[6] Li, Y., Zhang, L., & Liu, S. (2020). What happens for a ToF LiDAR in fog? IEEE Transactions on Intelligent Transportation Systems, 99, 1–12.

[7] Majer, F., Svoboda, J., & Novák, P. (2019). Learning to see through haze. En Proceedings of the European Conference on Mobile Robotics.

[8] McGuire, K. N., Kumar, V., & Michael, N. (2019). Minimal navigation solution for a swarm of tiny flying robots. Science Robotics, 4(35), eaau5660.

[9] Ruíz, P., Morales, R., & Herrera, J. (2014). El uso de imágenes LiDAR en Costa Rica. Revista Geológica de América Central, 51, 7–31.

[10] Wang, Y., Sun, Z., Liu, J., Shen, C., & Reid, I. (2019). Pseudo-LiDAR from Visual Depth Estimation. En Proceedings of the IEEE/CVPR (pp. 8445–8453).

[11] Xu, Z. C. B. (2018). Multi-Level Fusion Based 3D Object Detection From Monocular Images. En Proceedings of the IEEE/CVPR (pp. 2345–2353).

[12] You, Y. W., Wang, S., & Lu, H. (2020). Pseudo-LiDAR++. En International Conference on Learning Representations (ICLR).

[13] Zeng, W., Chen, X., & Shi, J. (2018). Inferring Point Clouds from Single Monocular Images. En Proceedings of the IEEE/CVPR.

Declaración sobre uso de Inteligencia Artificial (IA)

Para la revisión gramatical y ortográfica de este artículo, empleamos la herramienta de IA *ChatGPT*. Esta nos permitió identificar errores y mejorar la fluidez del texto. No obstante, realizamos una revisión final para garantizar que el artículo cumpliera con los estándares de calidad de la revista.