Navigation of UAVs in GPS-Denied Environments Using Computer Vision with Transformers

R. F. de A. S. Lima, B. R. B. Loureiro, L. Lanzellotti, D. N. G. Cavalcanti, J. V. G. Araujo, L. C. P. Macedo, and R. P. Fernandes ⓘ

Abstract— **This project tackles the challenge of enabling visual localization for UAVs in GPS-denied environments. We develop a similarity computation framework that compares aerial images using transformer-based embeddings. By leveraging DINOv2's rotation-robust features, the system processes image pairs to determine visual correspondence. This core module provides the foundation for future integration with georeferenced databases to support position estimation.**

Keywords— **DINOv2, unmanned aerial vehicles (UAV) Navigation, GPS-denied environments, transformers, image-based localization, Self-Supervised Learning, visual navigation.**

I. Introduction

Unmanned Aerial Vehicles (UAVs) are increasingly utilized across a diverse range of applications, including environmental monitoring, delivery services, and critical surveillance operations. The efficacy of most of these operations inherently relies on Global Positioning System (GPS) for precise navigation and localization. However, in specific challenging scenarios, such as dense forests or environments experiencing intentional signal blocking (jamming), GPS signals can become completely unavailable or severely unreliable. This presents a significant and critical challenge for the safe, continuous, and autonomous operation of UAVs.

To overcome this limitation, alternative navigation methods are needed, and one promising solution is visual navigation. In this project, we propose a proof-of-concept system that computes visual similarity between aerial images using robust embeddings. This foundational module enables future integration with georeferenced databases for position estimation. To accomplish this, we use computer vision techniques based on transformers, a type of deep learning model that has shown strong performance in image recognition tasks. Specifically, we leverage DINOv2, a self-supervised transformer model known for its ability to extract robust and rotation-invariant image embeddings. Our work delivers a visual similarity framework for aerial image matching in GPS-denied environments.

A. About Navigation of UAVs in GPS-Denied Environments

Unmanned Aerial Vehicles (UAVs) are vital for numerous applications, yet their core functionality largely depends on Global Positioning System (GPS) for navigation. This reliance poses a critical challenge in GPS-denied environments—locations where satellite signals are unavailable or

Rigel P. Fernandes, Department of Technology, Ibmec, Rio de Janeiro-RJ, email: rigel.fernandes@professores.ibmec.edu.br; Thiago Silva, Department of Technology, Ibmec, Rio de Janeiro-RJ, email: thiago.silva@professores.ibmec.edu.br.

intentionally blocked. Examples include dense urban areas, subterranean spaces, thick forests, or conflict zones. In these conditions, the lack of reliable GPS data leads to a severe loss of situational awareness, significantly increasing risks such as collisions, navigation errors, and mission failure. Therefore, developing robust autonomous navigation alternatives is essential for ensuring UAV operational viability and safety in such challenging scenarios.

B. Anticipated Challenges

Despite the promising capabilities of visual navigation, this approach presents several challenges that must be addressed for practical deployment. A primary concern is the inherent dependency on a comprehensive, pre-mapped georeferenced image database, which needs to be accurately updated and managed. Furthermore, optimizing the DINOv2 model for real-time inference poses challenges related to computational latency and power consumption. Environmental factors such as extreme lighting variations or adverse weather conditions could also potentially impact the robustness of visual feature extraction and matching.

C. Proposed Approach and Contributions

To overcome this limitation, alternative navigation methods are needed, and one promising solution is visual navigation. In this project, we propose a system that compares images captured with pre-stored satellite images to estimate the UAV's position in environments where GPS is not available. To accomplish this, we use computer vision techniques based on transformers, a type of deep learning model that has shown strong performance in image recognition tasks. Specifically, we leverage DINOv2, a self-supervised transformer model known for its ability to extract robust and rotation-invariant image embeddings. Our goal is to process and compare these images to support autonomous navigation and contribute to the development of more reliable UAV systems for GPS-denied scenarios.

II. System description and problems

A. Problem Context

Unmanned Aerial Vehicles (UAVs) are indispensable tools for a multitude of applications, yet their operational capabilities are fundamentally reliant on GPS for accurate navigation and localization. This dependency poses a critical challenge in GPS-denied environments—locations where satellite signals

are unavailable or intentionally blocked. Such conditions significantly hinder the safe and autonomous operation of UAVs, leading to severe limitations in their deployment. Current navigation paradigms, or the lack of robust alternatives, suffer from three key shortcomings:

- **Single Point of Failure**: The primary reliance on GPS means that operations are entirely crippled when satellite signals are lost, leading to an immediate breakdown in navigation capability.
- **Increased Operational Risks**: Without reliable positioning, UAVs face drastically heightened risks of collisions, navigation errors, and ultimately, mission failure in critical scenarios.
- **Restricted Operational Domains**: The inability to operate autonomously in GPS-denied environments severely limits the application of UAVs in crucial areas such as subterranean spaces or zones experiencing signal jamming.

B. Proposed System Pipeline

Our system processes pairs of satellite images to compute visual similarity using DINOv2's robust embeddings. The pipeline operates as follows:

- **Input**: Two RGB satellite images (Image (1) and Image (2)).
- **Processing**:
  1) Resize images to 224×224 pixels.
  2) Normalize pixel values using DINOv2's processor.
- **Feature Extraction**:
  1) Generate embeddings via DINOv2's Vision Transformer.
  2) Apply spatial average pooling.
  3) L2-normalize embeddings.
- **Rotation-Invariant Matching**:
  1) Rotate Image (2) at 180 degrees intervals (0 degree, 180 degrees).
  2) Compute cosine similarity for each rotation.
- **Decision**:
  1) Classify as "same location" if similarity > 0.87 (empirical threshold).

C. System Limitations

While our system leverages the robust capabilities of DINOv2 for visual navigation, certain limitations must be considered for its practical deployment and optimal performance:

- **Pairwise image comparison**: The current prototype focuses on pairwise image comparison. Future versions must address integration with georeferenced databases for coordinate inference and optimization for embedded hardware deployment.
- **Computational Demands for Onboard Deployment**: Achieving real-time inference using DINOv2 on resource-constrained embedded hardware is crucial for autonomous navigation.
- **Generalization to Extreme Environmental Conditions**: While DINOv2 shows strong robustness to variations in

illumination and generalization to diverse environments, its performance might still be affected by extreme and unseen environmental conditions, such as severe weather, very poor visibility (e.g., heavy fog, dust storms).

III. SOLUTION

This project develops a visual similarity framework for Unmanned Aerial Vehicles (UAVs) in GPS-denied environments. Its core functionality compares real-time aerial images against a georeferenced satellite image database to support position estimation.

Our system uses transformer-based computer vision models (DINOv2) to extract rotation-robust image embeddings. These embeddings serve as unique fingerprints, enabling similarity scoring between UAV-captured images and reference satellite imagery.

This solution aims to provide a reliable framework for visual localization, demonstrating a crucial step towards autonomous navigation. It contributes to the development of more resilient systems by demonstrating the effectiveness of DINOv2 in scenarios where traditional satellite guidance is not available, paving the way for future applications in a variety of challenging environments.

IV. RESULTS

Our initial results successfully demonstrate the core capability of our visual localization system: accurately computing image similarity using DINOv2's robust feature extraction. This involves preprocessing images, generating unique embeddings with DINOv2, and calculating their similarity via cosine distance to find the optimal rotational match. Future efforts should focus on integrating this visual localization core with broader navigation systems, optimizing its performance for real-time operation on embedded hardware, and managing the scalability of the georeferenced image database for practical deployment.

V. CONCLUSIONS

Although the project is still in its early stages, initial results from the developed system, leveraging DINOv2's robust image embeddings, demonstrate the significant potential for accurate visual localization in GPS-denied scenarios. The current prototype validates DINOv2's efficacy for rotation-robust image comparison. Next steps include integrating this module with a georeferenced image database to enable coordinate inference and optimizing inference. [1]

REFERENCES

[1] R. F. A. S. Lima, D. N. G. Cavalcanti, B. R. B. Loureiro, L. Lanzellotti, J. V. G. de Araújo, L. C. P. Macedo, and R. P. Fernandes, "Navigation of uavs in gps-denied environments using computer vision with transformers," in *XLII Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*. Natal, RN, Brazil: Sociedade Brasileira de Telecomunicações, September 29th, October 2nd 2025. [Online]. Available: http://dx.doi.org/10.14209/sbrt.2024.1571036315