MIT 6.S058: Drone-to-Satellite Image Alignment

Term Project Proposal

Spring 2026

1 Introduction

In high-stakes environments like "urban canyons" or natural disaster zones, standard GPS often fails because tall buildings reflect signals or ground infrastructure has been destroyed. To overcome this, we can give drones a "visual compass" that allows them to navigate by comparing what they see in real-time to a global satellite map. This technology has a wide range of applications - monitoring vulnerable ecosystems (like coral feefs), guiding life saving search-and-rescue drones through disaster-stricken ares, or smart infrastructure inspection.

The core challenge of this project is Image Alignment between two very different perspectives: the tilted, high-detail view from a drone and the flat, top-down view of a satellite map. To bridge this gap, we use a geometric "warp" (known as Planar Homography) to transform the drone's perspective so it looks like it was taken from directly above.

Once the perspectives match, the system uses Feature Extraction to identify unique landmarks—like the specific intersection of two roads or the footprint of a building—that appear in both images. By mathematically "locking" these live visual features onto the static satellite map, the drone can determine its exact coordinates without ever needing a satellite signal.

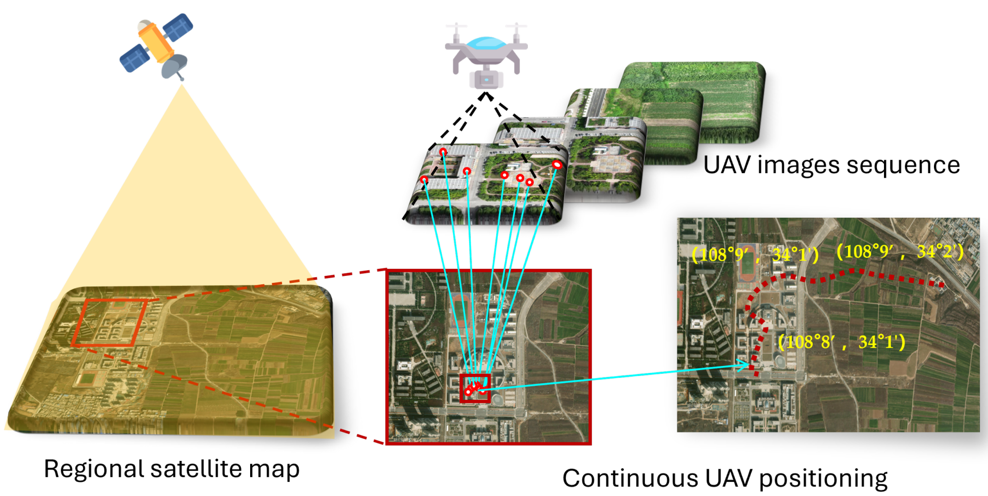

Figure 1: Satellite to Drone Imagery. Source: Scene-Adaptive UAV Visual Localization.

2 Problem Statement

Your goal is to align a "live" drone image to a static satellite reference. You must mathematically map pixels from one perspective to another and produce a composite image where the drone view is correctly overlaid on the map.

3 Methodology Options

3.1 Path A: Classical Feature-Based Alignment

- **Detection:** Identify stable image landmarks that remain detectable regardless of the differing resolutions. These keypoints serve as the geometric anchors required to compute a robust spatial mapping. See for example, the textbook's mention of local image detectors SIFT, SURF, and ORB [6]
- **Robust Estimation:** Implement a 3×3 Homography matrix and use RANSAC (see textbook) [3] to filter out outliers caused by transient objects like moving cars.

3.2 Path B: Structural & Edge Matching

- **Feature Engineering:** Would Canny edge detection or Hough Lines help align rigid man-made structures (buildings/roads) better than point-features? [1]
- **Metric:** Define and implement a "match quality" metric, such as Chamfer distance or cross-correlation.

3.3 Path C: Modern Learned Descriptors

- **Deep Features:** Use a pre-trained model to extract features that are invariant to the drastic lighting and seasonal changes between satellite maps and drone footage.

4 Resources

- **Dataset:** University-1652 [4] - Multi-view data for drone and satellite triplets.
- **Technical Blog:** Homography in Practice - A guide to the math of mapping planes.
- **Real-World Motivation:** Cornell: CV for Building Layouts.

5 Expansion Ideas

- Can you incorporate temporal consistency? If you have a video, can you use the previous frame's transformation to "prime" the search for the current frame?
- Semantic-Aware Masking: One of the biggest failure points in RANSAC is "distractor" movement from cars or swaying trees. Use a simple segmentation model to mask out non-static classes, forcing the alignment algorithm to focus exclusively on permanent structural features like road markings and building corners.
- If you have a drone, you can try out using your own data! Or try out with this drone video of MIT

[1] Read this wikipedia page for info on Canny edge detection and how it's implemented in this paper [5]. The first two steps involve 1) applying a gaussian filter and 2) finding the intensity gradient of the image—all concepts we have learned in class!

References

[1] Lowe, D. G. (2004). *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision.
Original Paper PDF.

[2] Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). *ORB: An Efficient Alternative to SIFT or SURF*. IEEE International Conference on Computer Vision (ICCV).
IEEE Xplore Link.

[3] Torralba, A., Isola, P., and Freeman, W. (2024). *Foundations of Computer Vision*, Chapter 41: Homographies. The MIT Press.
Foundations of Comupter Vision: Chaper 41.

[4] Zheng, Z., Wei, Y., and Yang, Y. (2020). *University-1652: A Multi-view Multi-source Benchmark for Drone-based Geo-localization*. ACM International Conference on Multimedia.
arXiv:2002.12186.

[5] D Liu, Y Wang, Z Tang, L Li, L Gao, "Automatic comic page image understanding based on edge segment analysis" in Document Recognition and Retrieval XXI, 2014

[6] Torralba, A., Isola, P., and Freeman, W. (2024). *Foundations of Computer Vision*, Chapter 44: Multiview Geometry and Structure from Motion. The MIT Press.
Foundations of Comupter Vision: Chaper 44.