Geo-Localization of Sequential Aerial Data

Team: Artem Krayevskiy, Khrystyna Mysak, Dmytro Khamula

Overview

Given:

- **A sequence of small UAV-captured images,**
- **One large satellite image with known geographic coordinates,**

This project focuses on the visual geo-localization task using sequential aerial imagery. The main objective is to determine the precise location of a series of small images captured by an Unmanned Aerial Vehicle (UAV) within a larger, geo-referenced satellite image.

Data

Example of a satellite image for scene num. 5

Example of a drone image for scene num. 5

Our dataset contains 6,742 drone images, 7 distinct sites 11 satellite maps.
Sites:

- Changjiang-20
- Changjiang-23
- Taizhou-1
- Taizhou-6
- Yunnan
- Zhuxi
- Donghuayuan
- Huzhou-3
- Huzhou-6
- Huailai
- Shandan

Scene 05

05_0395.JPG
Lat: 24.6525, Lon: 102.3413

05_0009.JPG
Lat: 24.6660, Lon: 102.3460

05_0212.JPG
Lat: 24.6592, Lon: 102.3549

Scene 06

06_0152.JPG
Lat: 32.3672, Lon: 109.6485

06_0045.JPG
Lat: 32.3681, Lon: 109.6394

06_0037.JPG
Lat: 32.3640, Lon: 109.6408

Example of a drone images for scene num. 5, 6

Example of a drone images for scene num. 3, 4

Scene 03

03_0545.JPG
Lat: 32.3124, Lon: 119.8641

03_0377.JPG
Lat: 32.3033, Lon: 119.8865

03_0555.JPG
Lat: 32.3072, Lon: 119.8722

Scene 04

04_0292.JPG
Lat: 32.2239, Lon: 119.9234

04_0731.JPG
Lat: 32.1656, Lon: 119.9478

04_0339.JPG
Lat: 32.1776, Lon: 119.9332

We reviewed several existing approaches, but most focused on a different objective (photo geolocation) whereas our goal is the geo-localization of sequential aerial data. As a result, we only adopted a few common applicable techniques from those methods, such as dividing large satellite images into smaller cropped sections.

- **PlaNet**
- **PIGEON**
- **Wide-Area Image Geolocalization**

Literature

Methods tried

Brute force **CNN/ViT** **SIFT+FLANN** **ALIKED+LightGlue**

Feature Pyramid **Image Stitching**

As the starting point we wanted to do the straightforward template matching: just simply take a small drone picture and do the template matching with a satellite one.

The problem here was that firstly, drone images which are 3k x 2k on the satellite image may take only 100 by 100 pixels or even less. Also, these images might be taken from another perspective.

So we tried to **resize the image to different shapes**, like 100 x 100, 200 x 200, 500 x 500, and also **rotated them from 0 to 360 degrees with step 15 degrees**.

CNN

We used feature extraction with a **resnet50** architecture, and then, using **FLANN** matcher, match it with some area on satellite image

ViT

Using this approach was the most promising for us, as visual ViTs, unlike traditional CNN-based models that primarily focus on local spatial features, are capable of capturing global contextual information through self-attention mechanisms. But, very unfortunately, ViT returned results which were a disaster and for a sequence of 9 images it returned the same 9 coordinates.

SIFT + FLANN

Example on a smaller satellite image (not in the dataset)

1. Extract keypoints and descriptors using **SIFT** from both the UAV crop and the entire satellite image
2. We then use **FLANN-based matching** to identify correspondences
3. Estimate the Homography matrix using **RANSAC**, which allows for robust geometric alignment and visualization.

Chose this approach as our baseline and decided to move forward

→

Example of keypoints detected and matched on a smaller satellite map

Number of matched keypoints on a larger satellite image (very bad)

Methods tried to fix the issue

- Extracting keypoints with **ALIKED** and match them with **LightGlue**
- Play with some resizing (using **Feature Pyramid idea** from the paper) and correctly classify the block
- Improve the **accuracy of identifying the first crop** by stitching the first three UAV images into a single panoramic image. This wider view is likely to contain **more diverse keypoints**, enabling better matching with the satellite image. The **stitching** is achieved by **estimating transformations** between images using **feature matching and RANSAC-based homography**.

Results & Future work

All in all, we experimented with various methods, but none significantly outperformed our basic pipeline based on **SIFT and FLANN**.

For evaluation, we used the haversine distance, and on large satellite images, our approach achieved an average error of around **0.7 km**.

On smaller satellite images, the localization was nearly perfect, if not entirely accurate.

If we had more time and resources, the next steps we would consider include:

- **Incorporating deep learning-based feature extractors, such as SuperPoint or D2-Net, to obtain more robust and invariant descriptors.**
- **Fine-tuning feature matching strategies, perhaps by integrating geometric verification beyond RANSAC or combining global and local features.**

Thank You

Team: Artem Krayevskiy, Khrystyna Mysak, Dmytro Khamula