VISUAL GEOLOCALIZATION FOR UAV USING SATELLITE IMAGERY

Under guidance of Dr. Ketan Rajawat

Presented By:
Vartika Gupta (180849)
Shubham Gupta (180748)

Motivation

Unmanned air vehicles (UAVs) are emerging into many sectors of society and have potential to significantly impact daily life.

UAVs are being increasingly used to solve civilian problems like disaster relief, conservation of natural habitat, fighting crime and terror.

UAVs are a part of many commercial applications like aerial surveillance, cargo transport, forestry and agriculture. They are an integral part of the military and defence systems as well.

They can play an important role in the future because in addition to the relatively low cost, they can un man aircraft to perform “dull, dirty, and dangerous” work.

Introduction

For navigation of UAVs , we ideally choose IMUs. But these IMUs will suffer with drift error. This drift error might be small at the start, but it will become large because of accumulation.

Usually, GPS signal is used for compensating this accumulated error and get a global pose estimate. However, for accurate geolocalization using GPS, the UAV must be able to receive a direct line of sight from four or more GPS satellites.The major disadvantage of a GPS-based system is that GPS signal is susceptible to degradation because of tall buildings and hostile conditions.

In the absence of GPS, we can use the camera sensors attached on the UAV for estimating the global pose of UAV and further navigation.

Problem Statement

For UAV navigation in GPS denied environment, we formulate a scheme for matching the UAV camera image with georeferenced satellite imagery, within less time in an efficient manner

1. See if it needs to be broken down into 2

Literature Review

The existing approaches can be divided into 3 broad categories:

1. **Particle Filters:**
   a. Richer Convolutional Features are extracted using SOTA filters followed by a CNN [1] model on the SAR image and aerial image for enhanced image matching [2]
2. **Kalman Filters (or Extended Kalman Filters):**
   a. Location classification using cross-view image matching[3]
   b. Approximate coordinates of the aerial image with the help of INS/GPS Kalman Filters on IMU sensor data [4]
   c. The registered image(Median filter -> Sobel Edge Detector -> Scaling -> Alignment)is matched with a small batch of geo-referenced images selected using approximate coordinates of the aerial image with the help of INS/GPS Kalman Filters on sensor data [5]

Literature Review

3. Least Square Method (Graph-based Methods):

- **a.** CNN network learns to predict relative pose change between two consecutive frames which is followed by a deep temporal CNN for local pose aggregation. Neural Graph Optimizer aggregates the information of the trajectory using an attention-based RNN [6]
- b. Orb-SLAM[7] : Same features for tracking and mapping increases efficiency and reduces time. Lesser outliers compared to PTAM. Uses bag of words algorithm for keyframe selection for initial place recognition followed by orb extraction and mapping.

Preprocessing Technique: Phase correlation based scanning of the Fourier Transform of filtered data for illumination-invariant image matching[8]

Dataset

We are using the dataset introduced in the paper [9], referred to as Aerial Cities. The UAV images are captured from Google Earth, and the satellite images are captured from Google Maps. The satellite images are always oriented towards the North. The UAV images have a random heading (angle varying between 0 to 360 degrees). The UAV images have been taken from an altitude between 100m and 200m above the ground. The satellite images have been taken from an altitude of 300m above the ground.

Train - Validation Split:

We have taken the UAV and satellite images of 6 cities (9865 data points). The data points of 5 cities (9065 data points) are used for training and validating the network in a 80-20 split and 1 city is used for testing (800 data points).

Preprocessing

1. Image flip(33% of the images), rotation by 10-30 degrees randomly (33% of the images) of SAR images to match the UAV image orientation and have a more detailed database
2. Image scaling (33% of the images) to adjust the altitude of the SAR images according to the UAV image.
3. Converted RGB images into HSV as this reduces the climate and daytime dependency of the model and reduces misclassification
4. Added and shuffled an equal number of non-matching pairs of images and labels to the dataset of matching pairs
5. The database gets doubled after flipping, rotating and zooming in the images (19730 data points). After introducing equal number of non-matching pairs, the size of the database becomes 38460 data points, that get divided into the training, validation and test data proportionately.
6. The images are converted into (128, 128, 3) normalized tensors using PyTorch transforms.
7. We use PyTorch Data Loaders to load data to the model dynamically, in small batches of size 32 because the large number of RGB images exceed the RAM limit.

ORB Feature Point Extraction

1. ORB point features of all the SAR and UAV image are passed as input for image matching of the UAV image and SAR images.
2. It improves the performance of the Resnet network.
3. Image feature point detection is done according to [10] where we first, select the pixel p in the image and assume its brightness is I and set a brightness threshold T. Then, take pixel p as the center, select 16 pixels on a circle with a radius r, and compare the gray value between pixel p and other pixels on the circle. If the brightness of consecutive N points on the selected circle is greater than $I+T$ or less than $I-T$, then pixel p can be considered as a feature point.
4. These feature points are marks on the images that signify solid lines of landmarks that help in image detection and matching.

Adaptive Residual Network

1. Backbone Resnet-50 network with 4 layers of [3, 4, 6, 3] residual blocks.
2. AveragePooling and MaxPooling layer followed by Normalization is used as a fusion network on a pre-trained Resnet-50 network.
3. Used to avoid degradation in performance in new environments with a different dataset.

Hyperparameters

- Mean Squared Error loss function
- Adam Optimizer
- Four layer Resnet architecture with [3, 4, 6, 3] residual blocks
- ReLu activation function
- 50 epochs

	Loss	**Accuracy**
Training		
Validation		
Test		

Plots

Conclusion

The model easily overfits on small dataset because less number of batches are passed into the heavily layered adaptive network.

Dual networks are superior to siamese networks when the data comes from 2 different distributions, here, the satellite and UAV images. Our model has the ability to correctly predict same scene or not 9 times out of 10.

Orb features are invariant to illumination, rotation and image scaling, hence improve the performance.

The examples on which the model fails are very confusing even to human eye. The model gets confused when:

1. the major portion of a UAV image is covered with a single structure
2. there is a lot of greenery or repeated city structures in the images

Proposed Future Work

1. Phase correlation value can be used as the metric for image matching to deal with old SAR images recorded in different weather conditions and times of the day compared to aerial image.
2. Sobel filters followed by a CNN filter can be used to extract only the major details of the images like terrains, buildings, etc and dynamic features which may change with time like vehicles are not extracted.
3. For navigation, using ORB-SLAM features to enhance the accuracy and make the code time efficient can be further explored as it is the most efficient and fastest, cutting-edge method for feature mapping in images.

Thank You