Google Map Oriented Visual Navigation for UAVs

Mo Shan[1] and Zhi Gao[2]
[1]Department of Electrical and Computer Engineering, University of California San Diego
[2]Temasek Laboratories, National University of Singapore

Motivation

1. GPS could be jammed or blocked.
2. Pose estimation by fusing inertial measurement unit (IMU) and optical flow suffers from drift.
3. UAVs that take off and land in different positions which may not revisit the same scene.

Objective

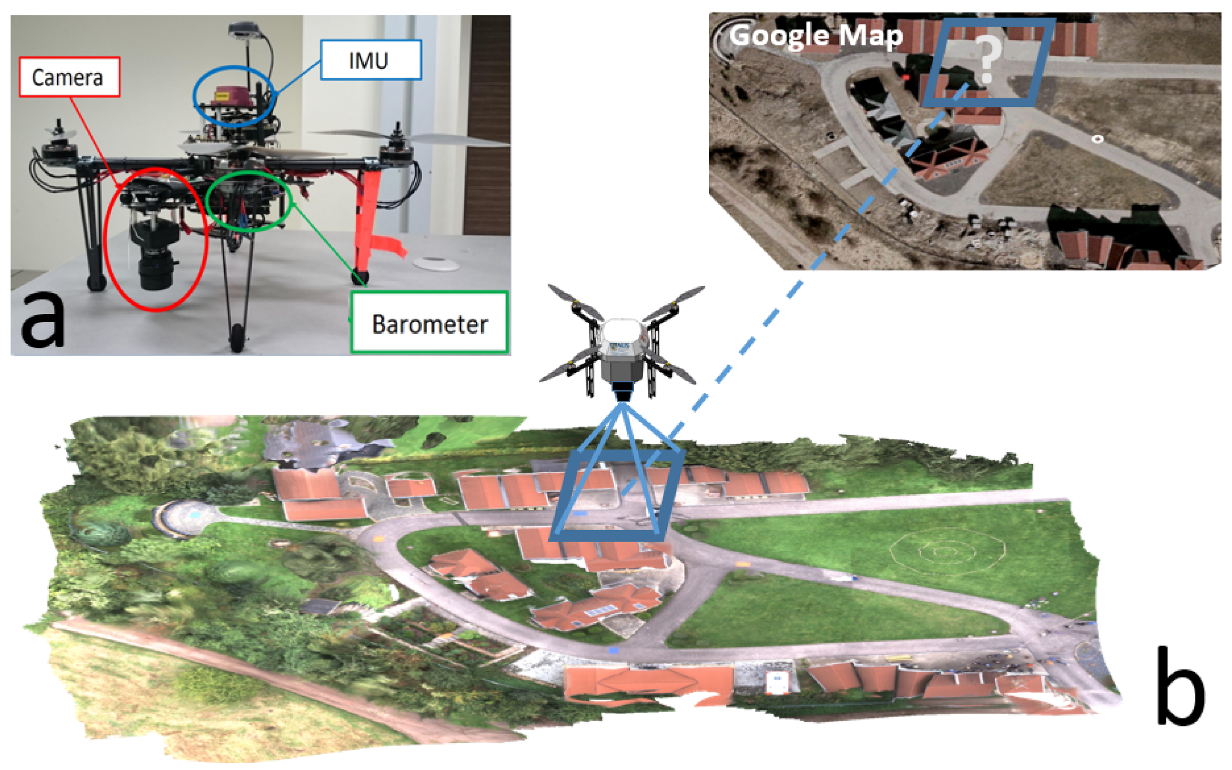

Figure 1: Overview. Sub-fig. a: The quadrotor platform. Sub-fig. b: Geo-referenced UAV navigation.

In this work, the localization of a UAV flying over the urban canyons is addressed. The quadrotor could reply on the on-board camera, IMU and the Google Map to estimate its position. Fig. 1 shows an overview of our approach.

Challenges

1. Significant scene changes due to difference in modality, viewpoint, weather, etc.
2. Lack of visible features in certain regions of low resolution map.
3. Large illumination variation for on-the-fly images.

Contributions

1. A robust visual navigation framework is explained in details. It uses a low resolution map for reference, which saves memory consumption and computational resources.
2. A simple yet effective navigation framework is proposed, which relies on correlation for initialization, Histogram of Oriented Gradients (HOG) features to describe the images, optical flow for motion prediction, and particle filters to avoid too many comparisons with sliding window.
3. To the best of our knowledge, this is the first time that low resolution Google Map is used for UAV navigation.

Geo-referenced navigation

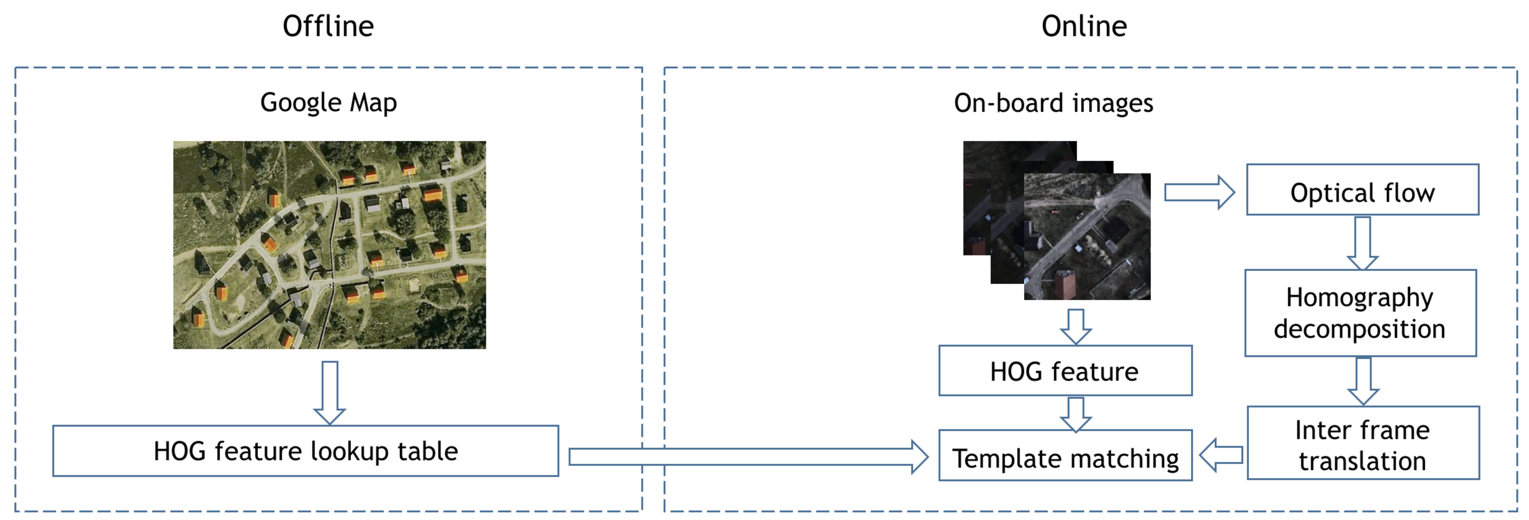

Figure 2: Pipeline of the proposed approach.

Fig. 2 shows the pipeline. HOG features for the map are computed offline. During onboard processing, we use global search to initialize the UAV position. Then for each frame, we track the pose by position prediction and image registration.

Global localization

After taking off, the UAV location is searched in the entire map for initialization. We correlate the current frame and the map as in Eq. 1, where F is the 2D Fourier transform of the input image: $F = \mathcal{F}(f)$, H is the transform of the map: $H = \mathcal{F}(h)$, $\odot$ denotes element wise multiplication and * indicates complex conjugate. As a result, transforming G into the spatial domain gives a confidence map of the location.

$$G = F \odot H^* \tag{1}$$

Pose tracking

Position prediction

We confine the matching around the predicted position. Shi-Tomasi corner detector and iterative Lucas-Kanade method with pyramids are used to obtain the optical flow. The inter-frame translation could be derived using Eq. 2 and Eq. 3, assuming the ground plane is flat:

$$\mathbf{H} = \mathbf{R} + \frac{1}{h}\mathbf{T}\mathbf{N}^\intercal \tag{2}$$

where $\mathbf{H}$ is the homography, $\mathbf{R}$ and $\mathbf{T}$ are the inter-frame rotation and translation, $\mathbf{N}$ is the normal vector of the ground plane, and h is the altitude. $\mathbf{R}$, $\mathbf{N}$, h are obtained from the on-board IMU and $\mathbf{T}$ can be calculated as

$$\mathbf{T} = h(\mathbf{H} - \mathbf{R})\mathbf{N} \tag{3}$$

Image description

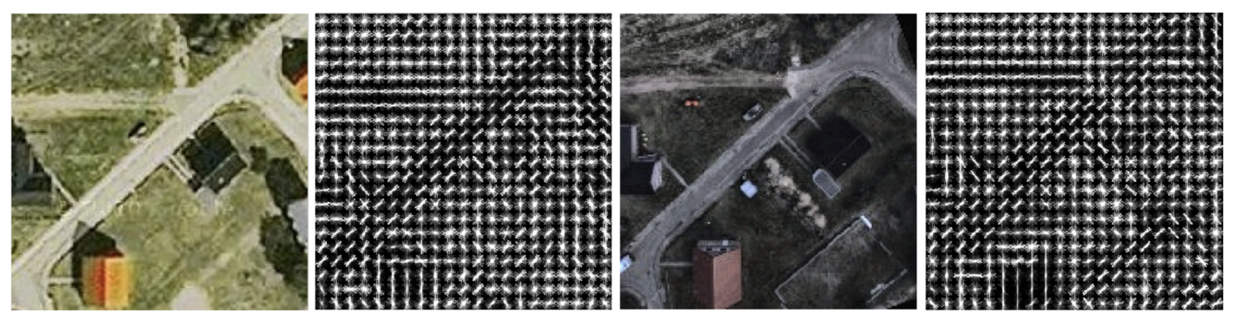

Figure 3: Visualization of HOG glyph.

We use HOG as an image descriptor to encode the gradient information in multi-modal images. The HOG glyph is visualized in Fig. 3. It is evident that the gradient patterns remain similar even though the on-board image undergoes photometric variations compared with the map. In particular, the structures of road and house are clearly preserved.

Coarse to fine image registration

Comparison of HOG features is time consuming, and thus we employ particle filters. There are N particles, and for each particle p, its properties include $\{x, y, H_x, H_y, w\}$, where (x, y) specify the top left pixel of the particle, (H_x, H_y) is the size of the subimage covered by the particle and w is the weight. Suppose each p predicts a location l, then the estimation is computed in Eq. 4.

$$E(l) = \sum_{i=1}^{N} w_i l_i \tag{4}$$

We use Gaussian distribution to normalize these distance values based on Eq. 5, where d_i is the distance between the two images under comparison, σ is the standard deviation, w_i is normalized using the sum of all weights to ensure that it lies in the range $[0, 1]$.

$$w_i = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-d_i^2}{2\sigma^2}\right) \tag{5}$$

For the coarse search, N particles are drawn randomly in a rectangular area, whose width and height are both s_c, with a large search interval Δ_c. The fine search is used if the minimum distance of coarse search is larger than a threshold τ_d, and it is confined in an smaller area with size s_f and search interval Δ_f. When the minimum distances in both coarse and fine search are above the threshold τ_d, i.e. an outlier, the predicted position is retained.

Results

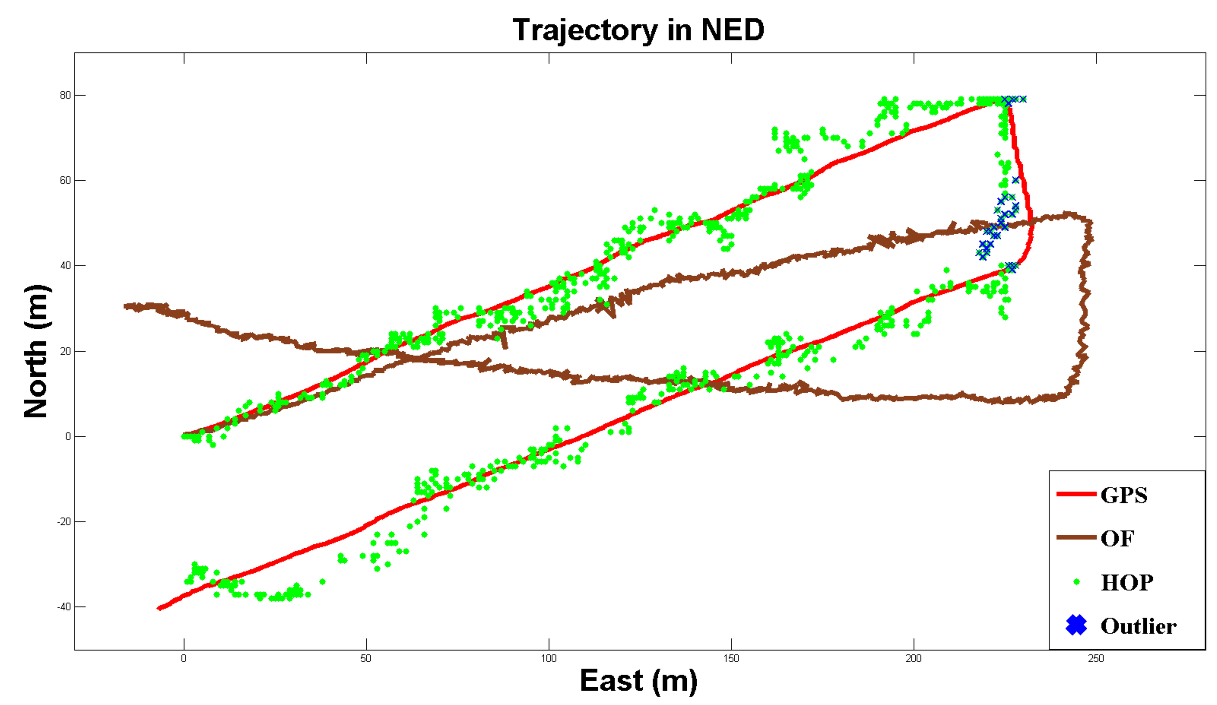

Figure 4: Path analysis comparing GPS (red line), OF for optical flow (brown line), and HOP for our method (green dots for reliable matches and blue crosses for outliers where OF is used).

As shown in Fig. 4, the proposed method is both accurate and reliable, because it takes advantage of not only the accuracy of HOG based localisation, but also the reliability of optical flow based position prediction. Compared to GPS, the root mean square error (RMSE) of our method is 6.773 m. The errors are quite small compared with a 169.188 m RMSE for the visual odometry based on optical flow alone.

Key insights

1. Low resolution Google Map could be used to provide prior information for localization.
2. HOG is an effective descriptor for multi-modal image registration.
3. Particle filter could be used in a coarse to fine scheme to increase image registration efficiency.