

Google Map Aided Visual Navigation for UAVs in GPS-denied Environment

Mo Shan^{1*}, Fei Wang¹, Feng Lin¹, Zhi Gao¹, Ya Z. Tang¹, Ben M. Chen²

Abstract—We propose a framework for Google Map aided UAV navigation in GPS-denied environment. Geo-referenced navigation provides drift-free localization and does not require loop closures. The UAV position is initialized via correlation, which is simple and efficient. We then use optical flow to predict its position in subsequent frames. During pose tracking, we obtain inter-frame translation either by motion field or homography decomposition, and we use HOG features for registration on Google Map. We employ particle filter to conduct a coarse to fine search to localize the UAV. Offline test using aerial images collected by our quadrotor platform shows promising results as our approach eliminates the drift in dead-reckoning, and the small localization error indicates the superiority of our approach as a supplement to GPS.

I. INTRODUCTION

Navigation of unmanned aerial vehicles (UAVs) in GPS-denied environment becomes increasingly critical. Since the UAVs often take off and land in different positions, it may not revisit the same scene. Thus, it may be difficult to detect loop closures as in simultaneous localization and mapping (SLAM). In addition, pose estimation via the fusion of inertial measurement unit (IMU) and optical flow (OF) suffers from drift [1]. To address these issues, we propose a geo-referenced localization framework, which is reliable, drift-free, and does not require loop closures.

We leverage on image registration to provide an absolute position in Google Map. Although its accessibility is appealing, and some relevant works have been reported, this task is quite demanding. Variation in scale, orientation, and illumination poses a great challenge to register the image captured by the onboard camera to the map. Furthermore, the scene changes between the onboard frame and the map is obvious because Google Map is not updated constantly.

To address these challenges, we rely on gradient patterns and use Histograms of Oriented Gradients (HOG) [2] for image registration. To expedite the matching process, we employ particle filter (PF) to avoid sliding window search. For efficiency, the search is confined around the UAV location predicted by OF.

Since our approach combines HOG, OF and PF, we coin the terms and name it HOP. In short, our contributions are summarized as follows. Firstly, we present a simple yet effective navigation framework, which relies on correlation for initialization, HOG features to describe the images and PF to reduce the amount of comparisons. Secondly, we propose an

OF based approach to compute inter-frame motion. To the best of our knowledge, this is the first time that low resolution Google Map and HOG are used for UAV navigation.

The rest of the paper is organized as follows: Section II presents a literature review; the detailed implementation of HOP is illustrated in Section III; Section IV contains experiments and analysis; following it are the conclusion and future research directions.

II. RELATED WORK

Various methods have been developed for UAV navigation to deal with GPS disruption. These works rely on different geographic information, including Geographic Information System (GIS), Google Earth, and Google Street View.

GIS data and its vector layers have been used to estimate UAV position in early works on geo-referencing. In [3], GIS data in the form of Ordnance Survey (OS) layers are divided into overlapping tiles, and inertial navigation system (INS) estimates which tile the UAV is located. The aerial image is rotated, scaled, and classified to form feature codes, which is compared with the OS data. Nevertheless, the training data may be inadequate to reflect the spectral signatures of all classes, especially under the presence of varying light conditions. Moreover, the classification, cross correlation and distance calculation are time consuming.

Images obtained via Google Earth have also been deployed for geo-referencing [4], [5]. In [4] the authors combine visual odometry and image registration to augment the navigation system. The Kalman filter is adopted to fuse the vision system with the INS. To compensate for the drift, an image registration technique realized by edge matching is developed. The registration is robust to change in scale, rotation and illumination to a certain extend. However, during the whole flight there are few successful matches. Therefore this method may not be suitable for long range flights. Another approach [5] that relies on Google Earth images involves classification of the scene. UAV images are segmented into superpixels and then classified as grass, asphalt and house. Circular regions are selected to construct the class histograms, which are rotation invariant. However, discarding rotation gives rise to the classification uncertainty. Consequently sometimes the drift in position estimation is not successfully removed. Moreover, the matching accuracy is poor in large homogeneous regions. In addition, the training requires labeling the reference map manually.

Recently, several approaches [6], [7] using Google Street View images for UAV localization have emerged. In [6], the aerial images are searched in the Google Street View

¹Temasek Laboratories, National University of Singapore, Singapore

²Department of Electrical & Computer Engineering, National University of Singapore, Singapore

*Email: shanmo@nus.edu.sg

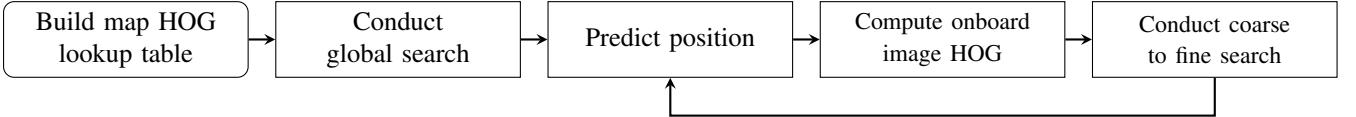


Fig. 1: Overview of HOP. The HOG features for the map are computed offline. During onboard processing, we use global search to initialize the UAV position. Then for each frame, we track the pose by position prediction and image registration.

database. To tackle with large viewpoint change, artificial views of the aerial images are generated and then Scale Invariant Feature Transform (SIFT) features are extracted. These features are compared against those extracted from ground images to find nearest neighbors. The outliers are removed via a histogram voting scheme and the good matches are verified by the Virtual Line Descriptor. Nevertheless, SIFT requires intensive computation and this approach does not use motion dynamics.

To summarize, our proposed approach mainly differs from the aforementioned ones in the following ways:

- 1) The easily accessible Google Map provides the geometric information for navigation, requiring less memory consumption compared with GIS and Google Street View.
- 2) The onboard sensors are utilized to obtain the rotation and scale of the frames, as well as the inter-frame translation.
- 3) It matches multi-modal images without feature detection and description. Instead, HOG is used holistically for image description and PF is employed to avoid sliding window search.

III. GEO-REFERENCED NAVIGATION

In this section the visual navigation framework will be investigated. An flowchart of HOP is given in Fig. 1, which provides an overview.

A. Global localization

After taking off, the UAV location is searched in the entire map for initialization. To avoid sliding window search, which is quite time consuming, we adopt the correlation filter proposed in [8]. In Eq. 1, F is the 2D Fourier transform of the input image, H is the transform of the filter, \odot denotes element wise multiplication and $*$ indicates complex conjugate. Because no training is available for detection, we correlate the current frame and the map. As a result, transforming G into the spatial domain gives a confidence map of the location.

$$G = F \odot H^* \quad (1)$$

Take the onboard image displayed in Fig. 2 for example, it corresponds to the rectangular region on the map. Its confidence map is shown in Fig. 3, from which it is evident that the correct location of the onboard image possesses the highest confidence. Although manual labeling may be more reliable, we still propose an autonomous global localization algorithm to make our framework complete.



Fig. 2: Onboard image at take off position, and its corresponding rectangular region in the map.

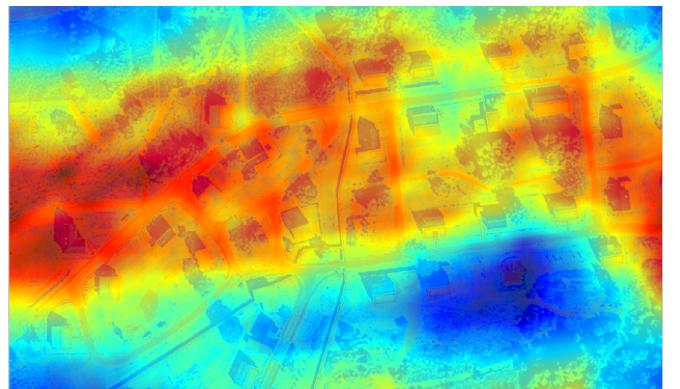


Fig. 3: The confidence map of the frame. Red indicates high confidence while blue indicates low confidence. The black area represents the highest confidence, which suggests that the UAV is at take off position. Best viewed in color.

B. Pose tracking

After initialization, the UAV position will be tracked based on local image registration. In this section, OF based motion estimation as well as HOG and PF based image registration will be introduced.

1) *Position prediction:* To narrow down the search, we make a rough guess and confine the matching around the predicted position by estimating the inter-frame motion. To obtain the motion, the points to be tracked are selected based on [9], and iterative Lucas-Kanade method with pyramids [10] is used to construct the OF fields. The inter-frame translation could be derived from two approaches based on [1], [11] respectively, both relying on supplementary information from the onboard avionic system and assuming the ground plane is flat.

Motion field: For an interest point P , its coordinates (x, y) in the camera frame and 3-D position $(X, Y, Z)^\top$ are related by $x = f \cdot \frac{X}{Z}$, $y = f \cdot \frac{Y}{Z}$ according to projective projection, where f is the focal length. Taking derivatives on both sides, we have $\dot{x} = v_x = f(\frac{\dot{X}}{Z} - \frac{X\dot{Z}}{Z^2})$, $\dot{y} = v_y = f(\frac{\dot{Y}}{Z} - \frac{Y\dot{Z}}{Z^2})$ where v_x, v_y are the OF.

Let the camera motion be expressed as a translation, $\mathbf{T} = (T_x, T_y, T_z)^\top$ and a rotation, $\Omega = (\omega_x, \omega_y, \omega_z)^\top$, the velocity of the feature point \mathbf{V} is defined by $\mathbf{V} = -\mathbf{T} - \Omega \times P$, whose explicit form is (2).

$$\begin{aligned}\dot{X} &= -T_x - \omega_y Z + \omega_z Y \\ \dot{Y} &= -T_y - \omega_z X + \omega_x Z \\ \dot{Z} &= -T_z - \omega_x Y + \omega_y X\end{aligned}\quad (2)$$

Therefore, v_x, v_y are related to the motion by (3).

$$\begin{aligned}v_x - (-\omega_y f + \omega_z y + \frac{\omega_x x y}{f} - \frac{\omega_y x^2}{f}) &= \frac{T_z x - T_x f}{Z} \\ v_y - (-\omega_x f + \omega_z x + \frac{\omega_y x y}{f} - \frac{\omega_x y^2}{f}) &= \frac{T_z y - T_y f}{Z}\end{aligned}\quad (3)$$

The rotational motion $(\omega_x, \omega_y, \omega_z)$ are read from IMU and feature depth Z is obtained by the barometer. Hence the terms on the left hand side are measurable. A linear equation set could be formulated for many feature points and the translation could be determined.

Homography decomposition: Homography describes the relationship between co-planar feature points in two images, from which the motion dynamics could be derived according to (4).

$$\mathbf{H} = \mathbf{R} + \frac{1}{h} \mathbf{T} \mathbf{N}^\top \quad (4)$$

The \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 onboard sensors and \mathbf{T} can be calculated as (5).

$$\mathbf{T} = h(\mathbf{H} - \mathbf{R})\mathbf{N} \quad (5)$$

2) **Image descriptor:** Unlike object detection, no training is available in navigation. Therefore, we use HOG in a holistic manner as an image descriptor to encode the gradient information in multi-modal images. The HOG glyph [12] is visualized in Fig. 4. It is evident that the gradient patterns remain similar even though the onboard image undergoes photometric variations compared with the map. In particular, the structures of road and house are clearly preserved.

During offline preparation phase, a lookup table is constructed to store the HOG features extracted at every pixel in the map. In this way, the HOG features for the map are retrieved from the table when registering images online to save computation time.

3) **Confined localization:** Comparison of HOG features is time consuming. Therefore, the traditional sliding window approach seems unfit as it demands more computational resources. Inspired by the tracking algorithms, we employ PF as in [13] to estimate the true position. Furthermore, in order

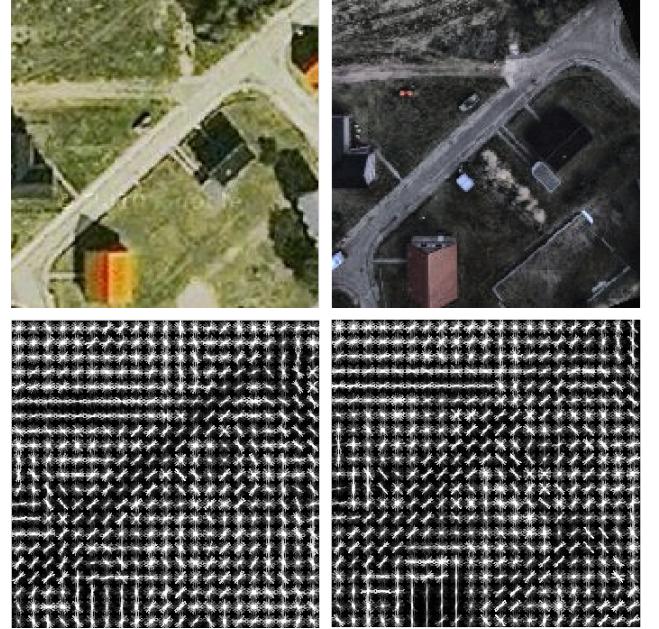


Fig. 4: Visualization of HOG histograms. First row: subimage of reference map and onboard image. Second row: HOG glyph. The gradient patterns for houses and roads are quite similar in HOG glyph.

to reduce the number of particles, we confine our search in the vicinity of the predicted position, adopting a coarse to fine procedure.

Particle filter: 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. The (x, y) is generated around the predicted position, while (H_x, H_y) equals to the size of the onboard image.

The optimal estimation of the posterior is the mean state of the particles. Suppose each p predicts a location l , then the estimated state is computed in Eq. 6.

$$E(l) = \sum_{i=1}^N w_i l_i \quad (6)$$

Based on the predicted state (x_p, y_p) of where the UAV could be in the next frame, we calculate the likelihood that UAV location (x_c, y_c) is actually at this location. After the particles are drawn, the subimages of the map located at the particles are compared with the current frame. To estimate the likelihood, we use Gaussian distribution to normalize these distance values based on Eq. 7, where d is the distance between the two images under comparison, σ is the standard deviation, \hat{w} is then normalized based on the sum of all weights to ensure that w is in the range $[0, 1]$.

$$\hat{w} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-d^2}{2\sigma^2}\right) \quad (7)$$

We do not use a dynamical model here to propagate the particles. Instead, we initialize the particles in every frame

using OF estimation, similar to [14], for we have to conduct coarse to fine search and the particle number changes from frame to frame.

Coarse to fine search: At the beginning, the particles are drawn around the take off position. Subsequently, OF provides translation between consecutive frames, and the predicted position is updated by accumulating the translation prior to image registration.

Around the predicted position, the search is conducted from coarse level to fine level to reduce the computational burden, similar to the coarse-to-fine procedure described in [15]. 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, on the other hand, is carried out in a smaller area with size s_f and search interval Δ_f . Different from [15], HOP relies mainly on coarse search which is often quite accurate. If the minimum distance of coarse search is larger than a threshold τ_d , then the match is considered invalid. Only when coarse search fails to produce valid match do we conduct fine search. Fine search still centers at the predicted position and the coarse search result is discarded.

When the minimum distances in both coarse and fine search are above the threshold τ_d , indicating that image registration result is unreliable, the predicted position by OF is retained as the current location. If the motion is too large, the UAV may conduct global search for re-initialization.

IV. EXPERIMENT

To evaluate the performance of HOP, we use the aerial images we have collected, which are displayed in the video accompanying the paper. The platform used to collect these images will be described, and then the pre-processing procedure as well as the parameter setting are elaborated. Next, we run experiments to analyse the effect of OF, and compare two outlier rejection schemes. We then compare HOP with visual odometry based on OF alone used as baseline.

A. Setup

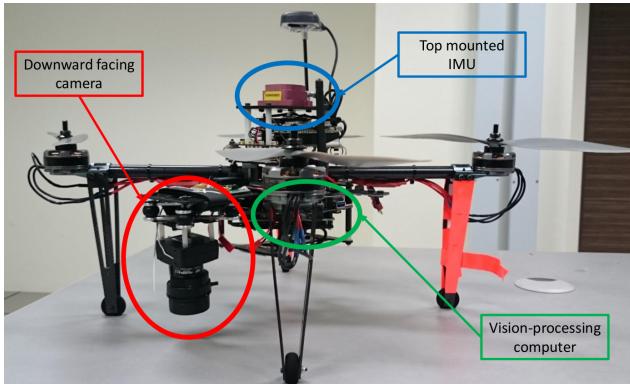


Fig. 5: Quadrotor platform used for image capture, whose onboard sensors include IMU, Mastermind, and a camera.

In order to test the performance of the proposed algorithm, data is collected onboard using a quadrotor platform shown

in Fig. 5, whose dimension is 35 cm in height and 86 cm in diagonal width with a maximum take-off weight of 3 kg. Its onboard sensors include an IG-500N attitude and heading reference system (AHRS) from SBG Systems and a downward-facing PointGrey Chameleon camera. An Ascending Technologies Mastermind computer is used to decode and log data in a 5 Hz rate. The flight test is carried out at Oostdorp⁺, the Netherlands, and the onboard images and IMU data collected are the actual fly-off data for the final round of the 2014 International Micro Air Vehicle Competition (IMAV 2014). The quadrotor flies at about 80 m above the ground and sweeps overhead the whole Oostdorp village. The speed is about 2 m/s and the total flight duration is about 3 min.

B. Preprocessing

The reference with size $w \times h$ is the Google Map of Oostdorp village, which corresponds roughly to a 300 × 150 m region. The resolution of the map is low, for about 3.15 pixels represent 1 m. The onboard frames are undistorted and then pre-processed to have the same orientation and scale as the reference map. First, the onboard frame is rotated by the yaw angle. Second, the frame is scaled to 3.15 pixels/m. The frames are cropped from the center with the same size $s_i \times s_i$.

C. Parameters

The most important parameters in HOP are N and s_c . More N increases the accuracy of the weighted center but demands more computational resources. Likewise, larger s_c ensures the matching is robust to jitter while smaller s_c reduces the time consumed. Hence, we trade off the robustness and efficiency when determining those parametric values. Regarding the sensitivity of HOP to these parameters, it is found that N should be larger than 40 to have sufficient particles to make a valid estimation. Meanwhile, s_c should be larger than 35 to account for the inaccuracy arises from OF.

In our experiment, the varied parameters used are set as follows. During preprocessing, $w \times h = 850 \times 500$, and $s_i = 180$. We use the HOG in OpenCV with cell size 32×32 , block size 64×64 , block stride 32×32 . For coarse to fine search, we set $N = 50$, $s_c = 40$, $\Delta_c = 4$, $s_f = 20$, $\Delta_f = 1$, $\sigma = 0.01$, $\tau_d = 0.75$.

D. Results

1) *Effect of prediction:* We compare the effect of OF based position prediction as shown in Fig. 6. UAV localization fails without OF, especially when the motion is large. The localization is only resumed when the UAV flies close to a previously seen place. By contrast, using OF overcomes the large motion by moving the search region to the predicted position, and hence loop closure is unnecessary.

2) *Rejection of outlier:* We compare two methods to reject outliers, namely minimum distance (MD) as well as Peak to Sidelobe Ratio (PSR) defined in [8]. PSR is computed

⁺Google Map location [52.142815, 5.843196]

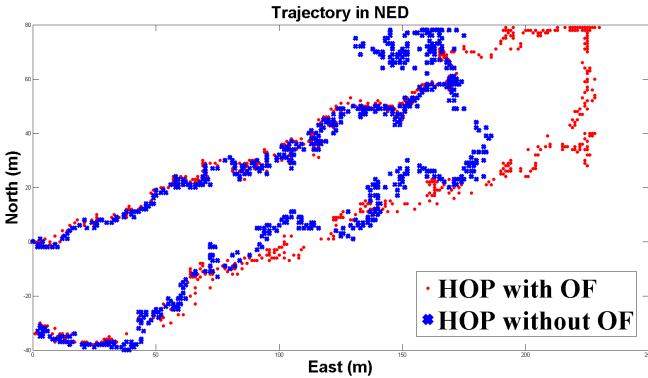


Fig. 6: Comparison of HOP and HOP without OF. Red dots represent HOP, while blue crosses represent HOP without OF. Without OF, HOP fails when there is large motion or unreliable match, while using OF handles those issues effectively. Best viewed in color.

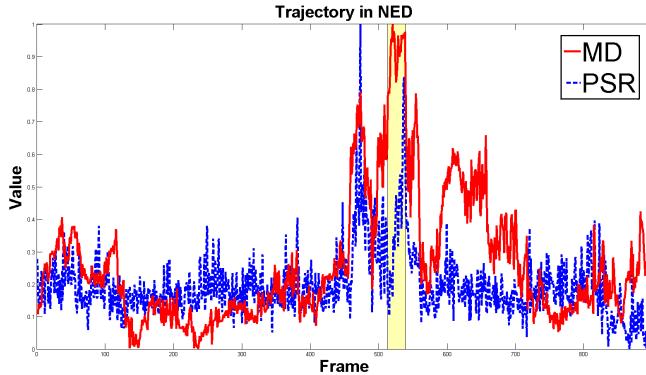


Fig. 7: Comparison of outlier rejection methods. Red line depicts MD, whereas blue line depicts PSR. The yellow area highlighted corresponds to the frames with large illumination change, when the match becomes unreliable. MD is significantly higher for unreliable match in comparison to PSR. Best viewed in color.

according to Eq. 8, where d_{min} is the minimum distance, and μ, σ are the mean and standard deviation of the distance for all particles in the search region, excluding a circle with 5 pixels radius around the minimum position.

$$\Theta = \frac{d_{min} - \mu}{\sigma} \quad (8)$$

The solid line in Fig. 7 is MD while the dashed line is PSR. Both MD and PSR are normalized to the range [0, 1]. MD peaks within the highlighted interval and attains large values, when the match becomes unreliable due to significant illumination change (refer to video). In contrast, PSR remains oscillating in that region. MD outperforms PSR in the sense that it indicates when the match is incorrect.

3) *Comparison with baseline:* The red line in Fig. 8 depicts the GPS ground truth, the brown line indicates the localisation from visual odometry based on OF alone as

baseline. The green dots represent the HOP output and the blue crosses are outliers. The sequence is challenging for image registration for three reasons. Firstly, the Google Map is not up to date, and the trees and buildings are missing in some region. Secondly, the map image only has low resolution, which may reduce the amount of visible gradient patterns. Moreover, the scene undergoes large illumination change (refer to video).

As shown in Fig. 8, HOP is both accurate and reliable, because it takes advantage of both the accuracy of HOG based localisation, and the reliability of OF based position prediction.

The dead-reckoning of OF gives poor results as the drift accumulates over time. On the other hand, the green dots follow the GPS closely, which corroborates the effectiveness of HOG based image registration. In comparison to ground truth, the root mean square error (RMSE) of HOP is 6.773 m. The errors are quite small compared with a 169.188 m RMSE for the visual odometry based on OF alone. In fact, the localisation accuracy of HOP is comparable with GPS, whose RMSE is 3 m.

Furthermore, the position prediction step in HOP deals with unreliable match effectively as well. When there is obvious illumination change around the second turn, the HOG based match produces low similarity, and the predicted position is closer to the ground truth. Hence the match is discarded as outlier.

Image registration failure constitutes 7% where position prediction is retained. The outliers mainly concentrate at two regions, where either there are few gradient patterns in the scene or has significant illumination change. We could design a flight path to avoid these homogeneous regions. Moreover, the oscillation of HOP is sometimes significant, mainly due to wind and jitter of UAV. A gimbal could be used to mitigate the oscillation.

E. Speed

HOP is implemented in C++ using OpenCV. It is not optimized for efficiency and runs at 15.625 Hz on average for each frame on a Intel i7 3.40 GHz processor. The current update rate of HOP is sufficient for the position measurement, since its output will be fused with onboard INS at 50 Hz [11]. Practically, the resulting trajectory is smooth as long as HOP is faster than 10 Hz.

V. CONCLUSIONS

This paper presents the first study of localization using HOG features in GPS-denied environment by registering aerial images and Google Map. The experiment using flight data shows that HOP could supplement GPS since its error is comparatively small. As the dataset is limited, our approach constitutes an initial benchmark, and we will make the onboard images and the reference map publicly available for research and comparison purposes.

For future research directions, we will install a gimbal to stabilize the camera against wind and vibration, and a thermal camera for navigation at night. Subsequently, we

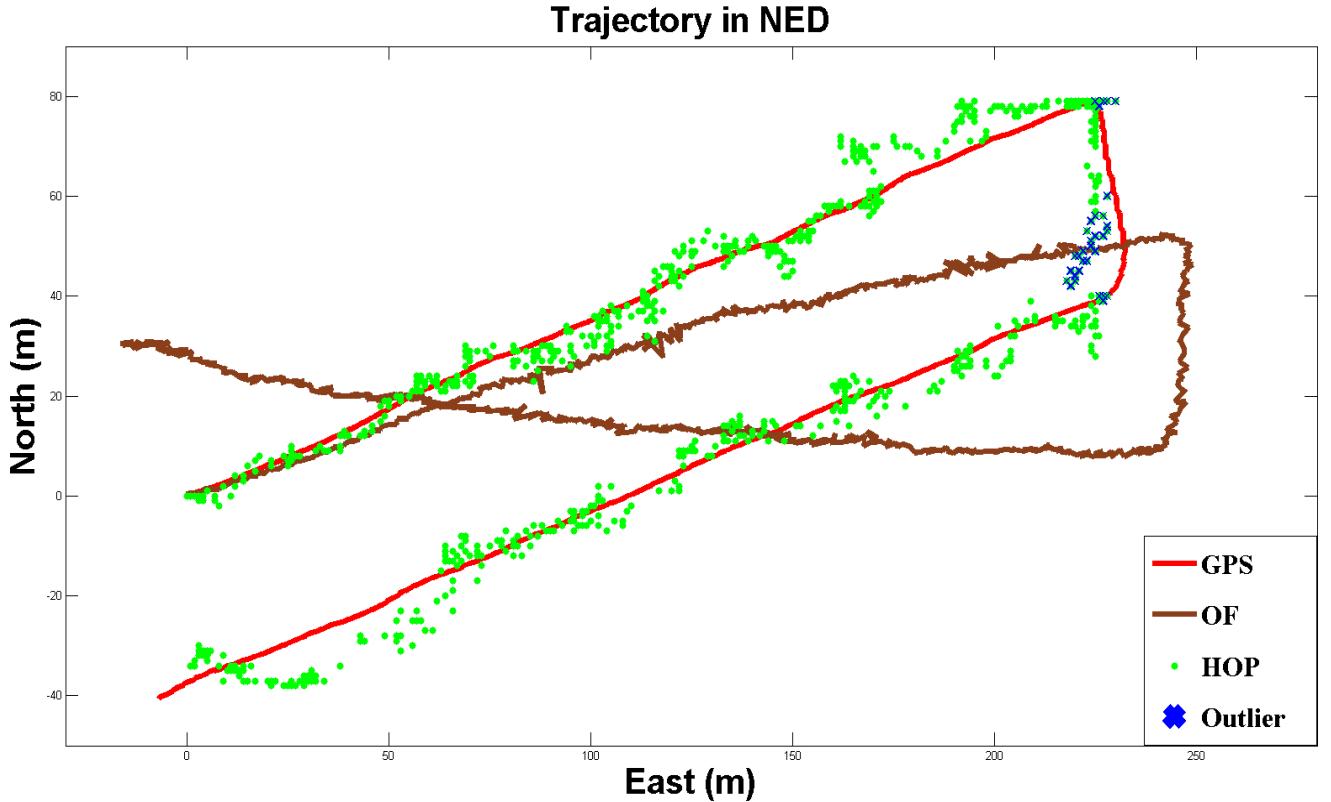


Fig. 8: Path analysis comparing GPS (red line), OF (brown line), and HOP (green dots for reliable matches and blue crosses for outliers where OF is used). HOP performs markedly better than the OF baseline. Best viewed in color.

will perform more evaluation on challenging environments, including day and night conditions. We will also use HOP to give state feedback to an actual system.

ACKNOWLEDGMENT

The authors would like to thank the members of NUS UAV Research Group for their kind support.

REFERENCES

- [1] D. Honegger, L. Meier, P. Tanskanen, and M. Pollefeys, “An open source and open hardware embedded metric optical flow cmos camera for indoor and outdoor applications,” in *Robotics and Automation (ICRA), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1736–1741.
- [2] N. Dalal and B. Triggs, “Histograms of oriented gradients for human detection,” in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, vol. 1. IEEE, 2005, pp. 886–893.
- [3] T. Patterson, S. McClean, P. Morrow, and G. Parr, “Utilizing geographic information system data for unmanned aerial vehicle position estimation,” in *2011 Canadian Conference on Computer and Robot Vision (CRV)*. IEEE, 2011, pp. 8–15.
- [4] G. Conte and P. Doherty, “An integrated uav navigation system based on aerial image matching,” in *Aerospace Conference, 2008 IEEE*. IEEE, 2008, pp. 1–10.
- [5] F. Lindsten, J. Callmer, H. Ohlsson, D. Tornqvist, T. Schon, and F. Gustafsson, “Geo-referencing for uav navigation using environmental classification,” in *Robotics and Automation (ICRA), 2010 IEEE International Conference on*. IEEE, 2010, pp. 1420–1425.
- [6] A. L. Majdik, Y. Albers-Schoenberg, and D. Scaramuzza, “Mav urban localization from google street view data,” in *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*. IEEE, 2013, pp. 3979–3986.
- [7] C. Le Barz, N. Thome, M. Cord, S. Herbin, and M. Sanfourche, “Global robot ego-localization combining image retrieval and hmm-based filtering,” in *6th Workshop on Planning, Perception and Navigation for Intelligent Vehicles*, 2014, p. 6.
- [8] D. S. Bolme, J. R. Beveridge, B. Draper, Y. M. Lui, et al., “Visual object tracking using adaptive correlation filters,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*. IEEE, 2010, pp. 2544–2550.
- [9] J. Shi and C. Tomasi, “Good features to track,” in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. IEEE, 1994, pp. 593–600.
- [10] J. yves Bouguet, “Pyramidal implementation of the lucas kanade feature tracker,” *Intel Corporation, Microprocessor Research Labs*, 2000.
- [11] S. Zhao, F. Lin, K. Peng, B. M. Chen, and T. H. Lee, “Homography-based vision-aided inertial navigation of uavs in unknown environments,” in *Proc. 2012 AIAA Guidance, Navigation, and Control Conference*, 2012.
- [12] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba, “Hoggles: Visualizing object detection features,” in *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE, 2013, pp. 1–8.
- [13] K. Nummiaro, E. Koller-Meier, and L. Van Gool, “An adaptive color-based particle filter,” *Image and vision computing*, vol. 21, no. 1, pp. 99–110, 2003.
- [14] A. Yao, D. Uebersax, J. Gall, and L. Van Gool, “Tracking people in broadcast sports,” in *Pattern Recognition*. Springer, 2010, pp. 151–161.
- [15] K. Zhang, L. Zhang, and M.-H. Yang, “Fast compressive tracking,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 36, no. 10, pp. 2002–2015, 2014.