GPS-independent localization framework for Aerial vehicles

An attempt at GPS-independent localization for Aerial vehicles using Machine learning, Machine Vision, and Monte Carlo localization

Vegard Bergsvik Øvstegård

September 15, 2020

UiO : Department of Informatics
University of Oslo

Contents

1 Introduction

In recent years, the popularity and use of Unmanned Aerial Vehicles (UAV) have increased drastically[1]. The usage and potential for both civilian and military applications are plural. From crop dusting and monitoring, infrastructure inspections to surveillance, and accident reporting. Drones have an increasingly larger presence and influence on our lives. With this in mind, it is very important that they and the systems they depend on work as intended.

Most classes of robots, in civilian and military applications, localization and navigation are fundamental capabilities. They are especially important for UAVs to execute complex operations. Inertial Navigation System (INS) and Global Positioning Systems (GPS) are often an integrated and crucial part of a UAV's navigation systems, despite their weaknesses.

State estimation from INS is rarely used alone and mostly aids position estimation when the GPS signal is unavailable or corrupt. However, due to hardware imperfections[2] that are very difficult to avoid, INS will drift over time.

Carrol [3] and Caballero[4] et al. state that the number of satellites and the quality of their respective signals play an important part in estimating a GPS receivers position. Few satellites or degraded signals will affect the estimation. Many factors affect GPS signals and estimation[5], those most relevant to Aerial Vehicles (AV) are for instance Signal Occlusion (SO). This happens when satellite signals are blocked due to buildings, bridges, trees, or other obstacles. When signals are reflected off buildings or walls, causing multipath propagation (MP). Jamming, Radio interference, or some Atmospheric conditions such as major solar storms can cause issues and corrupt signals. Satellite maintenance or maneuvers creating temporary gaps in coverage are also problematic.

As mentioned, UAVs are heavily dependent on GPS for navigation and localization. To obtain a robust positioning system, the many potential errors related to GPS has to be overcome. In recent years, active jamming of GPS signals has been claimed by the Norwegian government[6], resulting in a disruption of civilian flights. I.e having a GPS independent localization solution is of high importance and in some situations crucial. This motivates for developing alternate localization methods, able to work both alongside and independent of existing ones. Weight, space, size, and battery capacity are limited resources on UAV's, i.e there must be a cap on said factors and the upsides of the functionality must

outweigh the downsides of the added hardware.

In addition to GPS and INS, cameras are a very common sensor embedded on UAV platforms. With the continuing reduction in weight, price, and size compared to LIDAR or Laser Range Finder, the use of cameras has increased and is often regarded as a standard sensor. As images contain massive amounts of information about the environment, they are useful for many tasks. Vision-based localization systems are one of them and use only one or several embedded cameras on the UAV and a map of the environment. They do not rely on other external systems such as ground stations or satellites. Thus, a camera serves as a good candidate for a redundant localization solution or replacement when GPS fails.

Mantelli et al. [7] mention a possible approach and its challenges for the vision-based UAV localization problem. Using a downwards facing camera, providing orthophotos of the environment, and estimating the position of the images on an *a priori* known map that is based on aerial or satellite images. Most of the planet is already mapped using the aforementioned methods, there are many free and online sources providing maps like these such as Google™ Earth, Bing ™ Maps, and others. There are also increasingly more areas with detailed topography information from LIDARS and other sensors. Some of the quoted challenges with this problem are the update frequency and resolution of the maps. The images collected by the UAV might also have a significant difference compared to the *a priori* map. Illumination conditions, transient ground modifications caused by moving objects, weather conditions in particular rainfall and snow, but also long-term static modifications such as new roads or buildings.

Many works propose different types of maps and measurement models to overcome the challenges related to the UAV localization problem. Conserning this problem; maps are often 2D orthogonal maps or 3D point-clouds of the environment, and measurement models are functions that compute the similarity of the UAVs sensory data and a patch of the map. They do however come with caveats, some only work in specific scenarios where robustness falls out, and others have high computational cost or lacking precision. Hence a novel approach to improve such solutions is important to localize the UAV with high robustness and low power usage.

This essay proposes a somewhat novel strategy but is inspired by Mantelli et al. [7], Masselli et al.[8] and Nassar et al.[9]. A downward-facing camera and

a vision-based measurement model are used, and an extra step is added in an attempt to improve robustness and decrease the computational cost. The idea is to include an image segmentation network such as U-net[10], and use a very simple binary image descriptor on the segmented images from the camera and the *a priori* map. Robustness is induced by training the network to be invariant to some of the aforementioned challenges. E.g it should segment out long-lasting objects such as buildings despite some illumination conditions, weather, and seasonal changes such as snow. To estimate the UAV pose in 4 degrees of freedoms(DoF), the vision-based localization framework will apply the measurement model in a particle filter approach such as Monte Carlo Localization(MCL)[11]. In short, a segmented image of the UAV's view is compared to several random patches on the *a priori* segmented map, and the measurement model describes their similarity. With enough particles over time, the framework should find one that is similar enough to provide a likely position.

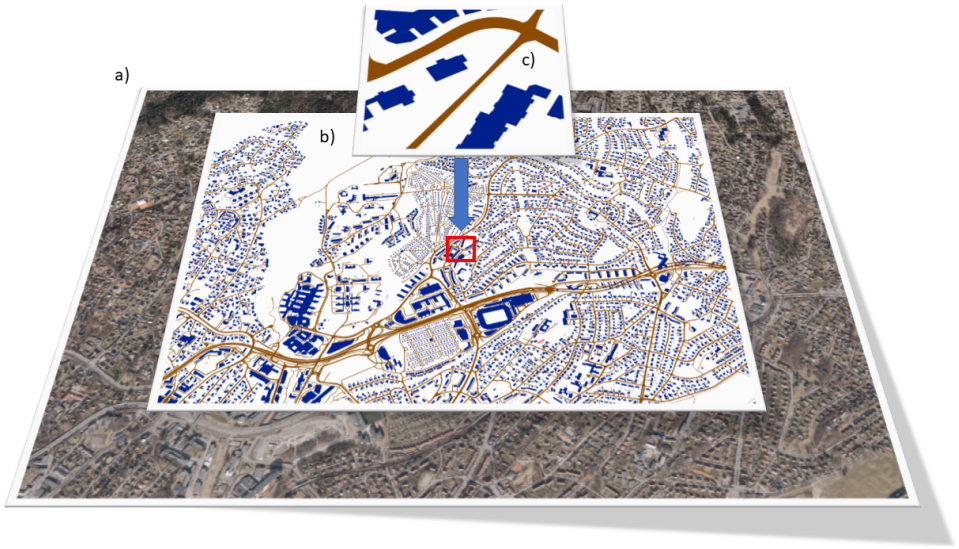

Figure 1: Global localization is made with comparing the segmented image captured by the UAV, **c)**, with the ground truth and *a priori* map **b)**, describing the environment **a)** with segmenting long-term objects such as buildings and roads.

As per Figure 1, the suggested framework compares the image observed and segmented by the robot, c), with a patch from the ground truth, b). As we are passing the observed image through a network, there might be a need to

pre-process the observed images. However, as the segmentation reduces the dimensionality of the images, the computational cost of comparing the observed image to a patch is lower. E.g the measurement model compares a few classes and not the different pixel values. This in turn resulting in higher robustness and hopefully lower computational cost. The approach is proposed as a redundant framework to estimate the UAV's pose. There are nevertheless some situations where the framework will have trouble producing a precise position as when passing over areas with little to no buildings or roads. There might also be errors when flying over homogeneous regions.

Prior simulations have shown that a simple binary measurement model used in an MCL approach is highly successful in localizing the UAV's pose, but this is given a perfect segmentation. However, with the stochastic nature of the MCL approach, a segmentation network of mediocre quality in accuracy and precision with some post-processing might produce decent results in localizing the UAV. Morphological operations such as dilation and erosion combined with shape matching using Hu moments [12] might improve localization as buildings more often than none have somewhat strict geometric shapes. I.e creating squares and rectangles out of the blobs that the network might produce. Nonetheless, the simulation is an indicator that the suggested framework does bear some merit.

2 Background and Related work

In recent years, advances in non-GPS localization of UAVs have been made, showing that this method has promise. While their implementation differs, MCL is often used. Many works propose vision-based solutions to the UAV localization problem using different approaches. They often deal with two different fundamental tasks, global localization and position tracking of which they will be divided. Furthermore, we show some works that use different descriptors for image matching in vision-based solutions, and also different approaches to image segmentation.

Global localization:
Masselli et al. [8] attempt UAV localization with a particle filter and terrain classification through feature extraction. Their solution provides global localization and an average error of $5.2m$ but is not proven to be robust against all

environmental changes, just some lighting, and seasonal changes. We believe that segmentation through Deep Learning will yield a much more accurate result that is more robust against environmental changes.
Mantelli et al. [7] propose a new localization strategy for a UAV equipped with a downward-facing camera, using a robust vision-based measurement model. The proposed measurement model computes the likelihood of the robot pose with the aid of an improved descriptor called abBRIEF [7], based on BRIEF [13]. The abBRIEF descriptor differs from BRIEF in two points: the color space used and the noise image reduction strategy. Their vision-based localization system applies the new measurement model in a Monte Carlo Localization (MCL) approach [11] that estimates the UAV pose in 4 degrees of freedom (DoF). In this paper the UAV is located within a short period, outperforming previous measurement models and yielding low errors, but is not proven to be robust against environmental changes like lighting and seasonal changes. We build upon this approach using a much simpler binary descriptor in combination with image segmentation.
Viswanathan et al. [14] demonstrate a working implementation of semantic segmentation with a Bayesian localization algorithm for ground vehicles across seasons, successfully localizing in satellite maps from summer, winter, and spring. Inherently, solving the localization problem is much harder for Unmanned Ground Vehicle (UGV) than for UAV, due to the drastic shift in perspective from the ground images to satellite map images. Although this paper also uses segmentation with LIDAR to locate roads. It gives merit to that invariance across seasons can be solved when using semantic segmentation and a particle filter. An important note is that their "winter" environment contained no snow, but this can be included when training the network.

Position tracking:
Nassar et al. [9] showing successful segmentation of satellite imagery using U-Net, but using a custom Semantic Shape Matching algorithm to establish the location in the satellite map. While the segmentation is largely successful, localization is sub-par and robustness against environmental changes is not proven. This framework also uses SIFT[15] Registration making the framework more computational heavy, and it does not inherently provide global localization.
Surber et al. [16] also presented an approach to localize a UAV locally, using the UAV's onboard visual-inertial sensor suite to first build a Reference Map of the UAV's workspace during a piloted reconnaissance flight. In subsequent flights over this area, the proposed framework combines keyframe-based visual-inertial

odometry with novel geometric image-based localization, to provide a real-time estimate of the UAV's pose with respect to the Reference Map paving the way towards completely automating repeated navigation in this workspace. The stability of the system is ensured by decoupling the local visual-inertial odometry from the global registration to the Reference Map, while GPS feeds are used as a weak prior for suggesting loop closures. The proposed framework is shown to outperform GPS localization significantly and diminishes drift effects via global image-based alignment for consistently robust performance.

Descriptors for image matching:
Zheng et al. [15] proposed an affine and rotation-invariant SIFT feature-based descriptor to perform matches between UAV and satellite images. This descriptor can vary the shape of the patch around a keypoint, becoming a robust descriptor with manageable computational complexity.
Calonder et al. [13] propose the use of binary strings as an efficient feature point descriptor, which we call BRIEF. We show that it is highly discriminative even when using relatively few bits and can be computed using simple intensity difference tests. Furthermore, the descriptor similarity can be evaluated using the Hamming distance, which is very efficient to compute, instead of the L2 norm as is usually done.

Segmentation:
Valada et al. [17] use a Deep Convoluted Neural Network with multiple modalities to achieve state-of-the-art performance on datasets with adverse environmental conditions, proving robust and reliable segmentation with acceptable inference time for mobile robotics. A 4.52 km trail through the Freiburg forest was driven autonomously using only the segmented images from the AdapNet with their Convoluted Mixture of Deep Experts (CMoDE), demonstrating the reliability and robustness claimed above.

3 Proposed method

The localization framework consists of two main building blocks:

- 1. A deep learning module is designed to extract certain semantic categories and produce synthetic images representing these categories.
- 2. A particle filter is designed for performing localization on said synthetic images.

The UAV captures an image, the deep learning module segments the image and feeds it into the MCL, which estimates the UAV position based on a segmented *a priori* known map. To lower the computational cost of the MCL, heading data from either the IMU or a possible Visual odometry system will also be feed into the MCL.

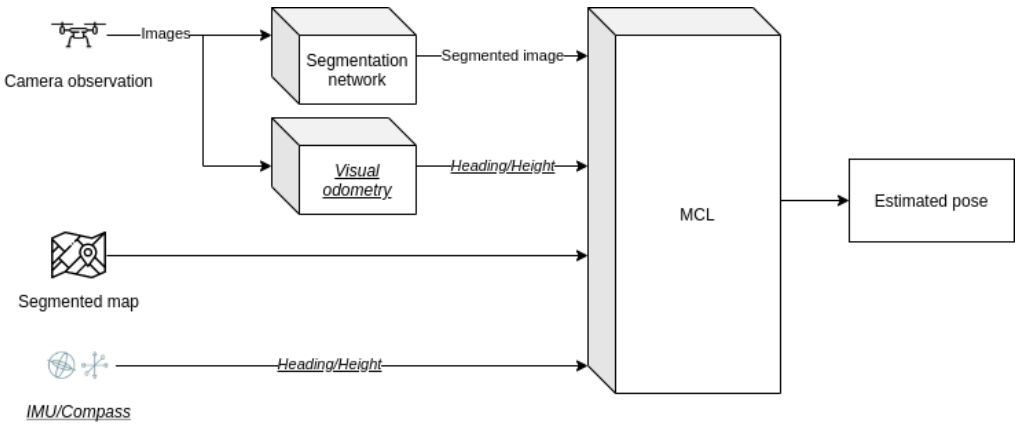

Figure 2: Flowchart over the proposed framework. Note the segmentation network requires training pre-flight, also to lessen the computational work of the MCL, heading, and possibly height is input from either an IMU, Compass, or Visual odometry.

3.1 Semantic segmentation

While MCL for UAVs with the use of satellite images has been proven to work with good results in previous works, the greatest shortcoming has been its lack of robustness against adverse environmental conditions such as differences in seasons, weather, and lighting. While classical digital image processing might improve robustness with regards to lighting conditions, deep learning has the

potential to make the MCL robust against most dynamic environmental factors, as it can learn features invariant to environmental changes and output a semantic representation of both satellite images and UAV images.

The first consideration for our implementation is U-net[10] as it is state-of-the-art on microscopic photography segmentation, who, like orthophotography, lacks three-dimensional features. Thus, we expect the U-Net to be performant on orthophotos as well, though other nets such as AdapNet [17], has shown to be better at segmenting multi-scale features, which could be important for a greater variance in drone height. There exist many variations of the U-net, with promising results that will also be taken into consideration.

3.1.1 Data set:

While many datasets for semantic segmentation of aerial and satellite images exist, very few contain the environment variances needed for our application. Thus another challenge is to modify or create a completely new dataset with the mentioned variances. As the Norwegian mapping organization, Kartverket provides detailed vectorized maps that can serve as ground truth for the datasets, see Figure 3, little work remains to create a large and functional dataset. Producing maps of different seasons and weather can easily be done with a drone and image mosaicking. Combining data from several sets will also aid in high variance possibly reducing overfitting and increasing generalization.

Figure 3: Example of Kartverkets vectorized images.

3.2 Monte Carlo Localization

The proposed framework utilizes segmented images from a downward-facing camera and localizes them in a segmented orthophoto that is used as a global map. Our approach applies an MCL algorithm, to locate the images which will be described in this section. The pixel comparison model has been simulated and provided outstanding results given a perfect segmentation. The model creates a set P of k pixels placed on the same location at both the particle and robot images. The intensities of each pixel x_i on both images are then compared and given a binary value depending on the pixels being equal or not. E.g if all pixels have the same intensity in both images, there is a high probability of the images being the same or similar depending on the number of pixels compared.

Figure 4: Example of pixel comparison. Here, a set P containing k = 4 pixels.

3.2.1 MCL Overview

Given a map of the environment, the goal of the algorithm is for the robot to determine its pose within the environment.

At every time t the algorithm takes as input the previous belief $X_{t-1} = \{x_{t-1}^{[1]}, x_{t-1}^{[2]}, \ldots, x_{t-1}^{[M]}\}$, an actuation command u_t, and data received from sensors z_t; and the algorithm outputs the new belief X_t. [18]

Algorithm MCL(X_{t-1}, u_t, z_t):
$\bar{X}_t = X_t = \emptyset$
for $m = 1$ to M:
$x_t^{[m]} =$ **motion_update**$(u_t, x_{t-1}^{[m]})$
$w_t^{[m]} =$ **sensor_update**$(z_t, x_t^{[m]})$
$\bar{X}_t = \bar{X}_t + \langle x_t^{[m]}, w_t^{[m]} \rangle$
endfor
for $m = 1$ to M:
draw $x_t^{[m]}$ from $\bar{X}_t$ with probability $\propto w_t^{[m]}$
$X_t = X_t + x_t^{[m]}$
endfor
return X_t

Figure 5: MCL algorithm

3.3 Optimizations and hardware

The variables and data used in the MCL algorithm and the measurement model, are inherently vectorized. With this in mind, optimizing the code execution to utilize heterogeneous multicore architectures may have a tremendous positive impact on execution time.

A development board such as nVIDIA's Jetson Xavier is a very good candidate for this framework. It provides an ARM-based processor and 8 Volta streaming multiprocessors and consumes a total of 15 Watt at max power draw. The unit is proven to be very efficient in AI usage as it also has 64 Tensor Cores.

References

[1] T. Murfin, "UAV Report: Growth trends & opportunities for 2019," *GPS world*. https://www.gpsworld.com/uav-report-growth-trends-opportunities-for-2019/.

[2] "INS drift," *Skybrary*. https://www.skybrary.aero/index.php/Inertial_Navigation_System_(INS).

[3] J. V. Carroll, "Vulnerability assessment of the u.s. Transportation infrastructure that relies on the global positioning system," *Journal of Navigation*, vol. 56, no. 2, pp. 185–193, 2003, [Online]. Available: www.scopus.com.

[4] F. Caballero, L. Merino, J. Ferruz, and A. Ollero, "Improving vision-based planar motion estimation for unmanned aerial vehicles through online mosaicing," in *Proceedings 2006 ieee international conference on robotics and automation, 2006. ICRA 2006.*, 2006, pp. 2860–2865.

[5] "GPS Accuracy," *GPS.gov*. https://www.gps.gov/systems/gps/performance/accuracy/.

[6] G. O'Dwyer, "Norway says it proved Russian GPS interference during NATO exercises," *GPS jamming*. https://www.defensenews.com/global/europe/2019/03/08/norway-alleges-signals-jamming-of-its-military-systems-by-russia/.

[7] M. Mantelli *et al.*, "A novel measurement model based on abBRIEF for global localization of a uav over satellite images," *Robotics and Autonomous Systems*, vol. 112, pp. 304–319, 2019, doi: https://doi.org/10.1016/j.robot.2018.12.006.

[8] J. Ginés, F. Martín, V. Matellán, F. J. Lera, and J. Balsa, "3D mapping for a reliable long-term navigation," in *ROBOT 2017: Third iberian robotics conference*, 2018, pp. 283–294.

[9] A. Nassar, K. Amer, R. ElHakim, and M. ElHelw, "A deep cnn-based framework for enhanced aerial imagery registration with applications to uav geolocalization," in *2018 ieee/cvf conference on computer vision and pattern recognition workshops (cvprw)*, 2018, pp. 1594–159410.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." 2015.

[11] J. Ginés, F. Martín, V. Matellán, F. J. Lera, and J. Balsa, "3D mapping for a reliable long-term navigation," in *ROBOT 2017: Third iberian robotics*

conference, 2018, pp. 283–294.

[12] Ming-Kuei Hu, "Visual pattern recognition by moment invariants," *IRE Transactions on Information Theory*, vol. 8, no. 2, pp. 179–187, 1962.

[13] M. Calonder, V. Lepetit, C. Strecha, and P. Fua, "BRIEF: Binary robust independent elementary features," in *Computer vision – eccv 2010*, 2010, pp. 778–792.

[14] A. Viswanathan, B. R. Pires, and D. Huber, "Vision-based robot localization across seasons and in remote locations," in *2016 ieee international conference on robotics and automation (icra)*, 2016, pp. 4815–4821.

[15] M. Zheng, C. Wu, D. Chen, and Z. Meng, "Rotation and affine-invariant sift descriptor for matching uav images with satellite images," in *Proceedings of 2014 ieee chinese guidance, navigation and control conference*, 2014, pp. 2624–2628.

[16] J. Surber, L. Teixeira, and M. Chli, "Robust visual-inertial localization with weak gps priors for repetitive uav flights," in *2017 ieee international conference on robotics and automation (icra)*, 2017, pp. 6300–6306.

[17] A. Valada, J. Vertens, A. Dhall, and W. Burgard, "AdapNet: Adaptive semantic segmentation in adverse environmental conditions," in *2017 ieee international conference on robotics and automation (icra)*, 2017, pp. 4644–4651.

[18] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics (intelligent robotics and autonomous agents)*. The MIT Press, 2005.