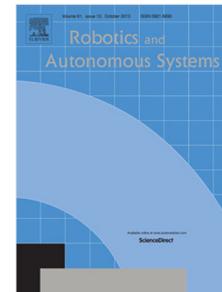


# Journal Pre-proof

■A review on absolute visual localization for UAV

Andy Couturier, Moulay A. Akhloufi

PII: S0921-8890(20)30506-6  
DOI: <https://doi.org/10.1016/j.robot.2020.103666>  
Reference: ROBOT 103666



To appear in: *Robotics and Autonomous Systems*

Received date: 23 September 2019

Revised date: 3 June 2020

Accepted date: 6 October 2020

Please cite this article as: A. Couturier and M.A. Akhloufi, A review on absolute visual localization for UAV, *Robotics and Autonomous Systems* (2020), doi: <https://doi.org/10.1016/j.robot.2020.103666>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier B.V.

# A review on absolute visual localization for UAV

Andy Couturier<sup>a</sup>, Moulay A. Akhloufi<sup>a,\*</sup>

<sup>a</sup>Perception, Robotics and Intelligent Machines Research Group (PRIME), Dept. of Computer Science, Université de Moncton, 18 Antonine-Maillet Ave, Moncton, NB, E1A 3E9, Canada

## Abstract

Research on unmanned aerial vehicles is growing as they are becoming less expensive and more available than before. The applications span a large number of areas and include border security, search and rescue, wildlife surveying, firefighting, precision agriculture, structure inspection, surveying and mapping, aerial photography, and recreational applications. These applications can require autonomous behavior which is only possible with a precise and robust self-localization. Until recently, the favored approach to localization was based on inertial sensors and global navigation satellite systems. However, global navigation satellite systems have multiple shortcomings related to long-distance radio communications (e.g. non-line-of-sight reception, multipath, spoofing). This motivated the development of new approaches to supplement or supplant satellite navigation. Absolute visual localization is one of the two main approaches to vision-based localization. The goal is to locate the current view of the UAV in a reference satellite map or georeferenced imagery from previous flights. Various approaches were proposed in this area and this paper reviews most of the literature in this field since 2015. The problematic at hand is analyzed and defined. Existing approaches are reviewed in 4 categories: template matching, feature points matching, deep learning and visual odometry.

**Keywords:** Absolute visual localization, UAV, navigation, satellite imagery, computer vision, deep learning

## 1. Introduction

The demand for unmanned aerial vehicles (UAV) is growing rapidly. In 2016, Goldman Sachs Research predicted that the global UAV market would be worth \$100 bn by 2020 [1–3]. The military segment being the largest one with \$70 bn followed by the consumer and the commercial segments which are expected to reach \$17 bn and \$13 bn respectively. The United States department of defense alone is asking \$9.39 bn in drone-related procurement, research and development, and construction funding in their 2019 budget presented to the U.S. congress [4]. This represents a 26% increase from the 2018 budget. As the UAV market grows, the range of applications for UAVs is also growing. First developed for the military [5], UAVs are now used for a wide variety of civil applications including: border security [6], search and rescue [7], wildlife surveying [8, 9], firefighting [10, 11], precision agriculture [12], structure inspection [13], surveying and mapping [14–16], aerial photography, and recreational applications. An important need for these applications is the requirement for autonomous or semi-autonomous operations. In most cases, humans only intervene to give high-level commands to the UAV navigation system. Autonomous navigation, in turn, requires a UAV to have a robust and reliable self-localization ability. Being a require-

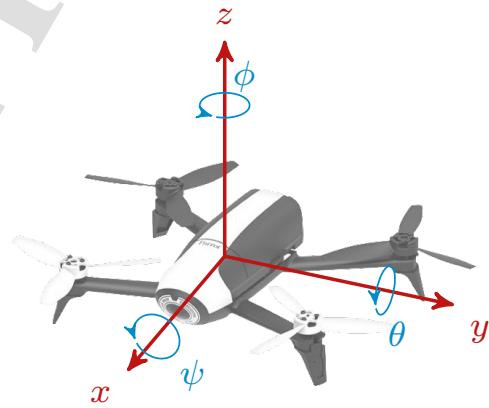


Figure 1: Six degrees of freedom localization.

ment for autonomous navigation, localization and pose estimation has been widely studied. Localization consists in estimating the pose of a UAV which is often represented by a 6 degrees of freedom (DoF) pose vector  $(x, y, z, \phi, \theta, \psi)$  as in figure 1. However, some UAV applications may require fewer parameters and employ simplified pose vectors such as the 4 DoF representation  $(x, y, z, \phi)$  and the 2 DoF representation  $(x, y)$ .

Early works in the field of UAV localization were using global navigation satellite systems (GNSS), such as the global positioning system (GPS), combined with an inertial navigation system (INS) inside a sensor fusion frame-

\*Corresponding author

Email addresses: eac6996@umanitoba.ca (Andy Couturier), moulay.akhloufi@umanitoba.ca (Moulay A. Akhloufi)

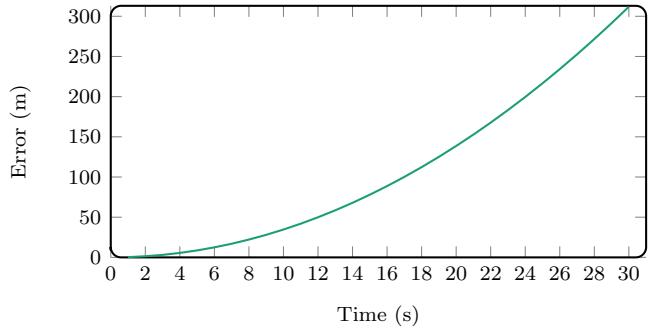


Figure 2: Simulation of the position estimation error with the Invensense MPU6050 accelerometer considering a systematic error of  $0.49 \text{ m/s}^2$  on both the x and the y axes.

work for pose estimation. These early works were using sequential Bayesian filtering techniques for GNSS-INS fusion such as Kalman filters [17, 18] and extended Kalman filters (EKF) [19]. GNSS-INS systems are mostly composed of a GPS, accelerometers, gyroscopes and magnetometers which are cost effective and lightweight sensors. Moreover, GNSS and INS have complementary characteristics. GNSS have good long-term accuracy and poor short-term accuracy, they therefore complement the INS that have the inverse characteristics. INS suffers from drift (error accumulation) due to the numerical integration required to convert acceleration measurements to displacement estimations. To illustrate this behavior we simulated a scenario where an accelerometer is used for a 2 DoF position estimation ( $x, y$ ). The simulated accelerometer make a systematic error of  $0.49 \text{ m/s}^2$  ( $50 \text{ mg}$ ) on both the x and the y axes. This is the typical error of a MPU6050 [20] a popular consumer grade inertial measurement unit (IMU) found in quadcopters such as the Parrot Bebop 2<sup>TM</sup> quadcopter [21]. Furthermore, we simulated readings at  $1 \text{ kHz}$  which is the maximum reporting frequency of the MPU6050. The results are illustrated in figure 2. The error is growing exponentially, reaching  $0.35 \text{ m}$  after 1 second,  $3.12 \text{ m}$  after 3 seconds and  $34.65 \text{ m}$  after 10 seconds. Note that real-life scenarios include other sources of error. For example, the gravity component needs to be removed from the accelerometer readings using gyroscope and magnetometer readings and thus, further error is induced.

Following the previous observations, it is clear that INS-only or IMU-only navigation, commonly known as dead reckoning, is unsuitable for UAV localization. This is why the global positioning information of GNSS is often added to bound the quickly growing error of the INS. GNSS can provide high quality readings but at a lower rate. Therefore, GNSS and INS are embedded, as previously stated, in a data fusion framework to correct the INS error when GNSS readings are available. Unfortunately, while this approach is interesting and elegant, it suffers from serious reliability issues. As GNSS works by measuring the time-of-flight of a radio signal [22], any obstacle

between a UAV and the emitting satellites can be problematic. Reception issues have been studied in indoor environments [23, 24] and in multiple outdoor environments such as urban canyons [25–31] and forests [32–38]. Signal unavailability is only one type of issue encountered but some other issues are a lot more insidious. Intentional [39] and unintentional [31, 40, 41] interference can degrade the signal and induce errors without completely rendering it unavailable. For example, multipath reception is a type of unintentional interference that occurs when one or multiple reflected signals are interfering with the real signal [31, 42]. This effect can be visualized in figure 3a. Non-line-of-sight (NLOS) reception is a phenomenon where the real signal is blocked and the GNSS receive a reflection of this same signal. The induced error in the position estimation is typically in tens of meters but can be more than a kilometer in some urban environments with tall buildings [43]. An illustration of that effect is shown in figure 3b. When this type of issue arises, the GNSS-INS solution can degrade to an INS-only solution with catastrophic ramifications. Furthermore, other types of problems exist such as signal spoofing which is a threat especially for defense and security applications. Signal spoofing, as illustrated in figure 4, is a security vulnerability that allows an attacker to take control of a UAV by making it think it is somewhere else through signal usurpation [44]. This type of attack can be extremely problematic in case of armed conflict [45]. As a result of these issues emanating from radio communications, a lot of research has been pursued in the fields of GNSS-denied and GNSS-degraded navigation. Vision-based and vision-aided localization has been the most predominant solutions to replace or supplement GNSS-INS fusion in recent years. The main challenges of such approaches reside in the vast amount of data to analyze and into the complexity of interpreting this data. This paper review recent works in this field and draw a portrait of the current state-of-the-art. Related works include surveys on computer vision approaches for UAVs [46, 47], vision-based navigation for UAVs [48], UAV navigation in GPS-denied environment [49] and even narrower subjects such as stereo-vision navigation [50]. However, these works have not put their focus entirely on the aspect of localization. In fact, these works alone may lead to a diluted view of the current state of the field of localization which is buried in other reviewed aspects of navigation such as path planning. We also note that recent literature on localization has not been yet reviewed. For example, the most up-to-date of the aforementioned works, [49], is only reviewing works from 2007 to 2015 and since then, many works have been published in this area. Furthermore, only relative visual localization (RVL) approaches such as visual odometry (VO) and simultaneous localization and mapping (SLAM) have been reviewed. Notwithstanding the lack of reviews on absolute visual localization (AVL), we have been able to find a work by Xu *et al.* [51] that touched on this subject. This work seems to be the only review of the literature in this area and is also the

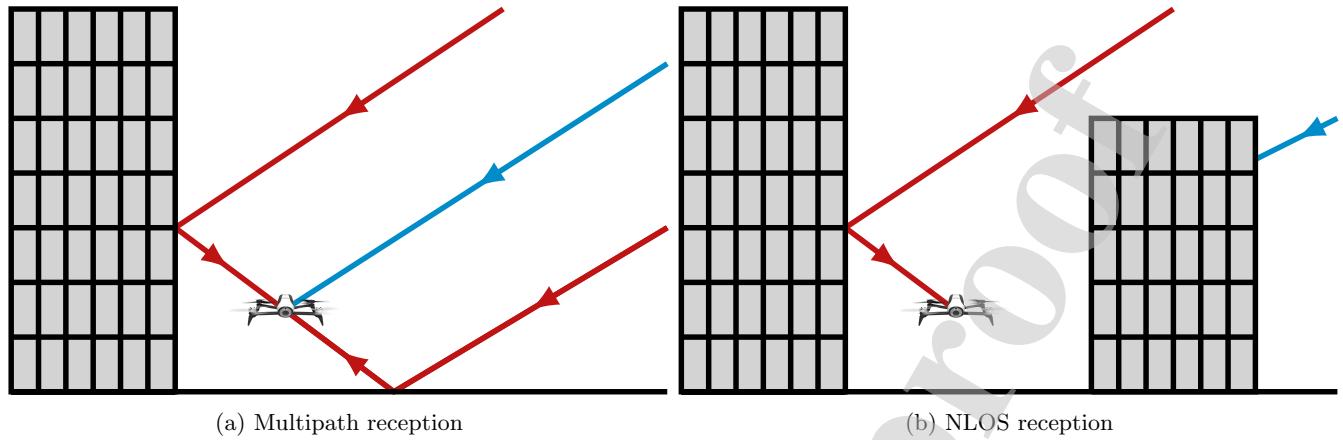
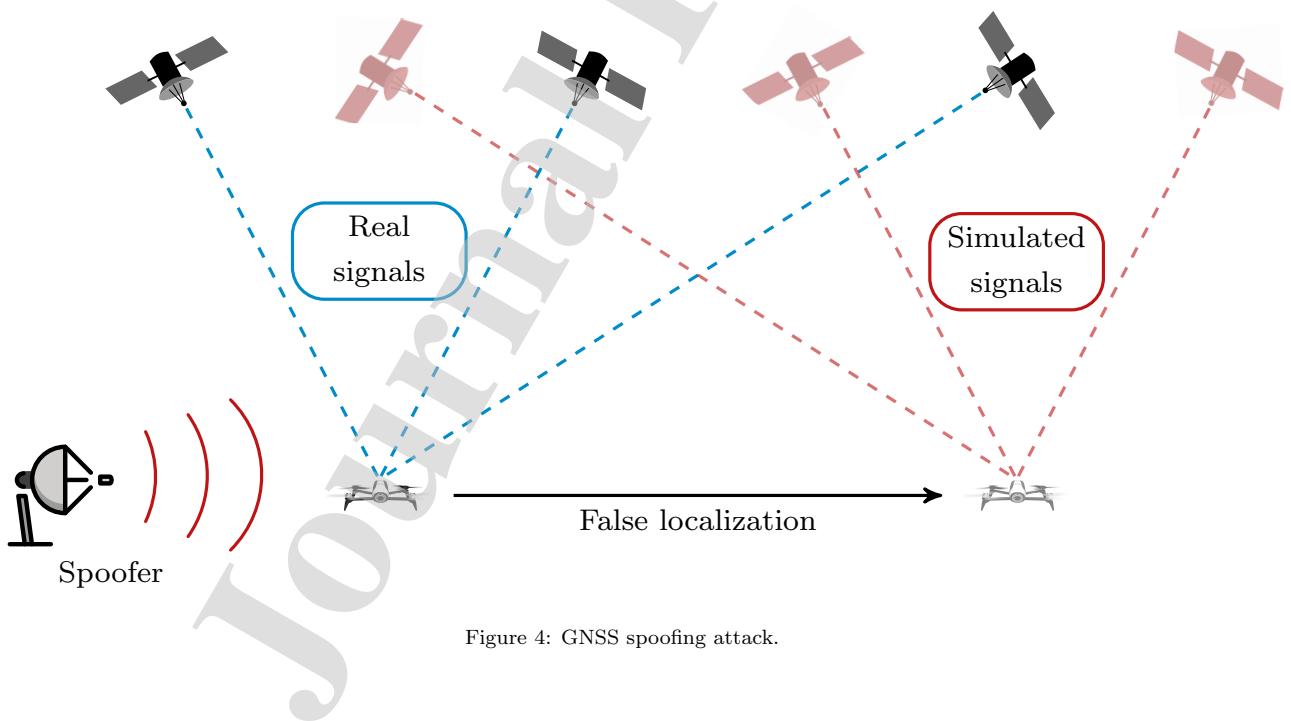


Figure 3: Examples of common GNSS reception issues. The blue lines represent the real signal and the red lines represent reflected signals.



closest work to ours. The authors presented techniques useful for AVL such as traditional and deep feature extraction methods and efficient feature representations. Works closely related to AVL were discussed such as visual loop closure for SLAM, content-based remote sensing image retrieval, UAV images to satellite images registration, place recognition and image geolocalization. The approach of the authors was to explore promising techniques applicable to AVL. Therefore, the authors reviewed works that could be related to AVL but did not specifically review AVL solutions with the exception of a couple of them. The paper had a focus on localization with large-scale reference maps and deep learning approaches for feature extraction. This focus explains why the authors did not review current AVL literature as almost all existing approaches assume the prior knowledge of the starting position or an approximation removing the need for localization in large-scale reference maps. Moreover, the current state-of-the-art in AVL is not yet using deep features and is instead using traditional feature extraction methods such as BRIDGE [52].

Following these observations, we propose a review covering the literature from 2015 and onward with a special focus on current absolute visual localization techniques for UAVs. This work has a very different take on the subject than previous works in this area. The key contributions of this paper are the following:

- A first review on absolute visual localization techniques for UAVs;
- A comparison of the approaches based on multiple variables and using tables for better visualization and comparison;
- A discussion of the reviewed works, their issues, and the challenges remaining in the field;
- An organized portrait of the future of the field through a list of possible research directions.

## 2. Relative Visual Localization

The field of vision-based UAV localization comprises two main approaches: relative visual localization (RVL) and absolute visual localization (AVL). These two approaches are also sometimes called frame to frame localization and frame to reference localization, respectively. Relative visual localization includes popular methods such as visual odometry (VO) and simultaneous localization and mapping (SLAM). To better understand the necessity and the usefulness of AVL, we must first look at these prominent RVL approaches.

VO is a technique that compares the current frame observed by the UAV with the previous frame to analyze the differences in egomotion. This is usually performed with optical flow (OF) analysis [53, 54]. The new pose estimation is obtained by adding an estimated difference pose vector to the previous pose estimation. Therefore,

VO is only using the current and the previous observation for each position estimation.

SLAM, on the other hand, is built around the concept of location recognition and map building. A map of the environment is constructed as the autonomous vehicle is exploring the environment [55]. SLAM is using concepts such as keyframes and bundle adjustment (BA) [56, 57] to take advantage of multiple previous observations. The core principle of SLAM is to jointly estimate landmark positions and the pose of the vehicle. A key characteristic of SLAM is the ability to recognize previously visited locations (i.e. loop closure) and to adjust the pose estimation and the map in accordance with this observation. Therefore, SLAM have the ability to use multiple previous observations to localize by using two different mechanisms. However, while SLAM is being able to use more information than VO, the current estimation still relies on a series of previous estimations. Moreover, in the absence of a possibility for loop closure (e.g. in large environment), the solution can quickly degrade and lose its benefits over VO.

The core issue with RVL is error accumulation, commonly known as drift over time. Drift is a byproduct of the recursive use of estimations to provide new estimations. The issue is trivial, if the current estimation relies on the previous ones, the error made in the prior estimations will impact the accuracy of the current estimation. This fundamental characteristic of RVL has sparked a large body of research aimed at mitigating the drift influence in these approaches. For example, inertial data from INS have been integrated tightly with VO to form very precise visual inertial odometry (VIO) systems [58–60] that are powering most of today's consumer UAVs. However, while being invaluable contributions, these approaches don't address the underlying source of error accumulation. In some applications, such as defense and security, immunity to drift is a highly desired characteristic. The UAVs employed in such applications are often high endurance UAVs. Some of these UAV can even stay in operation for up to one week at a time and cover areas as large as 725,197 km<sup>2</sup> [61]. In such applications, having known and constant error bounds can be of absolute necessity and this is precisely where AVL excels.

## 3. Absolute Visual Localization

AVL is an approach to localization that is radically different from RVL. The main advantage of AVL is its inherent immunity to drift over time. The method is using previously collected data (reference data) to localize the UAV. This data is assumed to be precisely georeferenced prior to its use for localization. The reference data may be composed of aerial imagery in the form of a loose set of images or aerial imagery stitched together to form a mosaic. These aerial images are most of the time orthorectified satellite imagery [62]. Nowadays, this data is more easily available than ever with the growing choice of freely available cartographic systems like Google Earth™ [63].

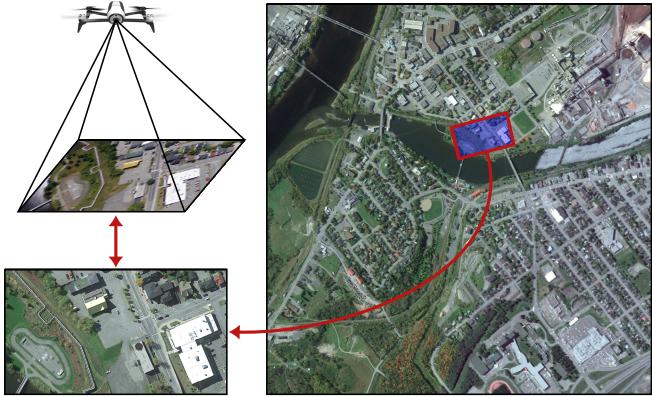


Figure 5: Absolute visual localization performed by matching UAV data with a reference satellite map.

and geographic information systems (GIS), like ArcGIS<sup>TM</sup> [64]. Another alternative is to use imagery collected from previous flights and georeferenced using an onboard GNSS [65–68]. This approach requires the GNSS to be reliable during the data collection but subsequent localization does not require GNSS. This type of data is also easily available with the widespread availability and the affordability of consumer quadcopters. The main goal of AVL is to match or register the current view of the UAV against a visual memory built from the aforementioned data in order to perform localization. The immunity to drift is achieved by ensuring a complete independence between the estimations. A high-level illustration of the AVL task is presented in figure 5. The main challenges related to this approach are the large amount of data required and the difficulty in matching or registering images from different sources. The images are often taken by different sensors, at different times, with different levels of illumination and different resolutions. A single reference map or reference image set can contain images from many sources. UAV imagery can also be significantly different from the reference images especially when these images are orthorectified and post-processed. Works in the field of AVL have addressed these challenges in numerous ways. Some authors addressed aspects related to search space reduction while others didn't. Nevertheless, all the reviewed works include a form or another of image matching or registration. The next sections will be discussing the 4 classes of image matching techniques found in the AVL literature: template matching, feature points matching, deep learning matching, and visual odometry matching. A visual representation of this organization is available in figure 6.

### 3.1. Template matching

Template matching is also known as direct or dense matching in the field of image matching and image registration [69, 70]. Multiple approaches using template matching has been proposed in the AVL literature. The task of localization can be redefined as a template matching problem where the current view of the UAV is used as a

template that is searched in a reference map. This type of approach uses an image patch comparison operator such as the sum of squared differences (SSD) to compare two image patches and obtain a measure of similarity. Authors that proposed such approaches for AVL have integrated the patch comparison function in search space reduction techniques. This is necessary as the main drawback of template matching is the high computational cost of the similarity measure. Recursive Bayesian filtering [71] and numerical optimization [72] have been used to efficiently retrieve the location of the template in the reference map. The following section reviews such approaches.

Dalen *et al.* [71] defined a technique to estimate the absolute position of a UAV using normalized cross-correlation (NCC) [73]. NCC was used as the probability density function (PDF) embedded in a particle filter framework [74]. It is used to perform dense image alignment between images captured by a UAV and a global reference map created from Bing Maps<sup>TM</sup> [75]. The authors assume that a measure of altitude is available to allow the scaling of the UAV images prior to the template matching. The aim of this work is to supplement an underlying SLAM-based navigation system [76] with an absolute position estimation when available. NCC is applied in a window around each particle during the measurement update step of the particle filtering process. After the particle filtering, the mean and variance are computed from the resulting PDF estimation defined by the distribution of the particles. For the sake of robustness, these results are only considered when a variance threshold is met and when successive estimates have been sufficiently close to each other. Once the conditions are met, the position estimate and the variance obtained are implemented into the measurement update step of an extended Kalman filter (EKF) [19] present in the underlying SLAM navigation system. Experiments have been performed outdoor with a GTMax helicopter which is a Yamaha RMax helicopter [77] with custom software. The environment of operation was approximately 90 m × 100 m. The position error between the position estimation of a differential GPS and the map alignment solution embedded in the SLAM solution was on average 3.6 m with a maximum of 12.5 m.

Yol *et al.* [72] used a dense approach based on mutual information (MI) as a similarity measure. MI is an information theory measure of the statistical dependency between two signals. In the context of UAV localization, MI is employed to measure the amount of information shared by two images [78, 79]. MI is more computationally expensive than SSD or NCC but is also more robust to local and global differences between the images being compared. The authors used a global reference map constructed from a mosaic of georeferenced images obtained from Google Earth<sup>TM</sup> [63] for their experiments. An assumption of a planar ground and of a UAV image plane parallel to it is made to allow for the use of the sRt model. The model is defined by the following equation:

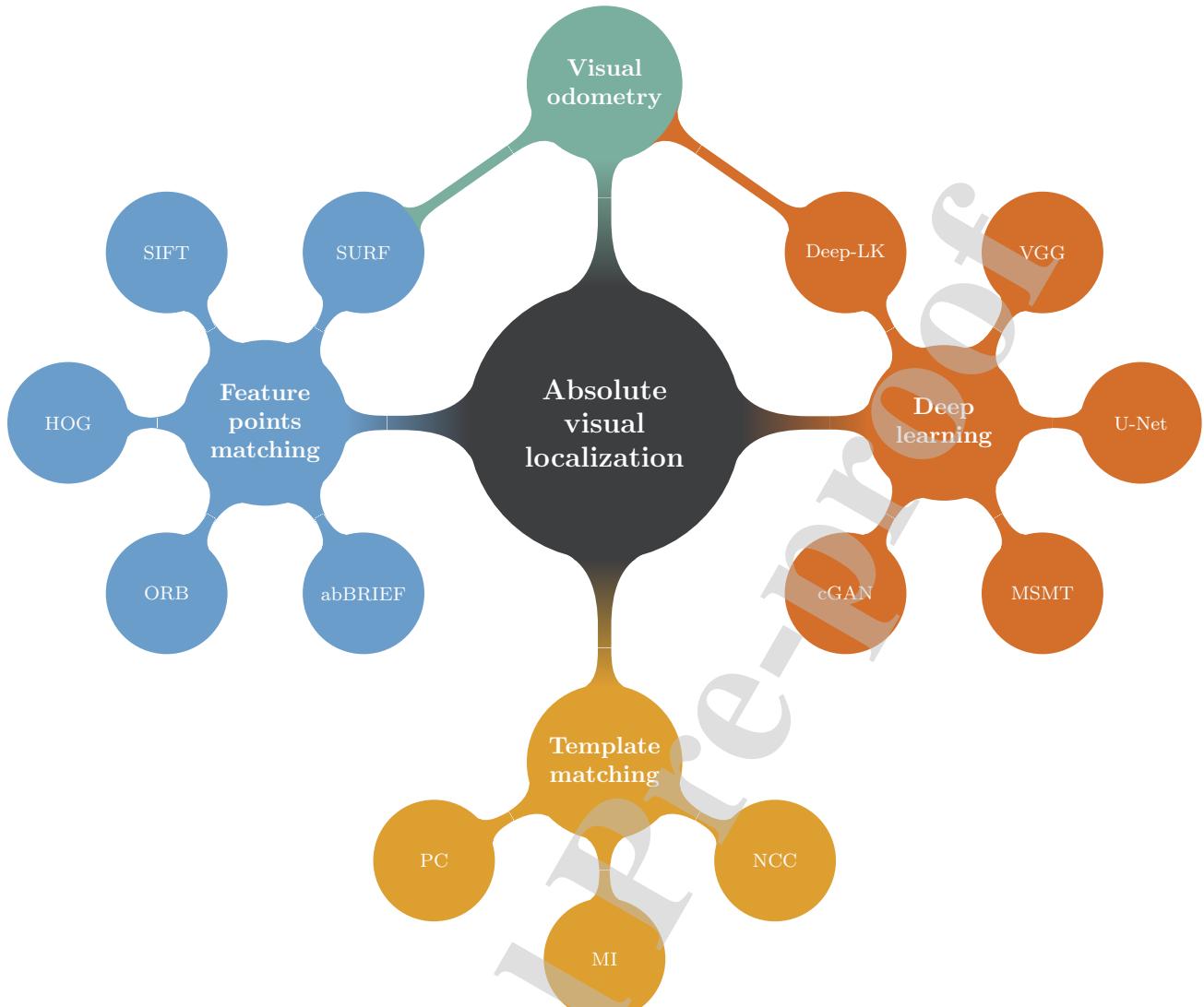


Figure 6: Organization and conceptual relations of the reviewed methods.

$$w(\mathbf{x}_t, \mu) = s\mathbf{R}_{2d}\mathbf{x}_t + \mathbf{t}_{2d}$$

It's a simple model that account only for a scale factor ( $s$ ), a 2D rotation ( $\mathbf{R}_{2d}$ ) and a 2D translation ( $\mathbf{t}_{2d}$ ). This is in opposition to homography models that also account for the roll and the pitch of the UAV which in this case would be an over-parameterization under the authors' initial assumptions. Newton's optimization method with a second order Taylor expansion is used to find the maximum of the MI function of a UAV image, warped in the  $s\mathbf{R}\mathbf{t}$  model, over the global reference map. Real-life experiments have been conducted with a UAV equipped with a camera mounted on a brushless gimbal. The flight was performed at an altitude of 150 m over a 695 m trajectory in a small outdoor environment. The results show a RMSE of 6.56 m in latitude, 8.02 m in longitude and 7.44 m in altitude.

Wan *et al.* [80] developed a localization solution built

around illumination invariant Phase Correlation (PC) [81]. PC is a template matching or image alignment approach based on the Fourier shift property. The Fourier shift property indicates that a translation shift between two similar images generates a linear phase difference in the Fourier frequency domain. Using PC as an image patch similarity measure requires to assume that the two images are radiometrically similar. The authors first hypothesized that this assumption would be broken if the two images have been taken with different illumination conditions. In fact, in previous research they had determined that the PC matrix  $Q(u, v)$  of such images can be decomposed into two independent PC matrices [80]:

$$Q(u, v) = Q_1(u, v) \times Q_2(u, v)$$

Where  $Q_1$  is an illumination impact matrix that accounts for the difference caused by illumination and  $Q_2(u, v) =$

$e^{i(au+bv)}$  is accounting for the translation. This decomposition implies that illumination invariant template matching can be achieved by eliminating the impact of  $Q_1$  on  $Q$  by extracting  $Q_2$ . The authors assessed the impact of  $Q_1$  in a simulation and under the assumption that a terrain surface can be approximated by a Lambertian surface and considering the sun's azimuth and zenith angle variations. During this experiment, they ended up empirically demonstrating that the Fourier fringe density and orientation remain unchanged. In practice, the fringes in  $Q$  are blurred in comparison to the ones that would be observable in  $Q_2$  however, the shift remain the same as in  $Q_2$ . Therefore, the conclusion is that the PC algorithm is fundamentally insensitive to illumination variations caused by the sun's position. Following this first contribution, the authors proposed a PC-based AVL solution using satellite imagery. The UAV's images are first rectified with the data from an inertial measurement unit (IMU) and a method described in [82] to obtain an approximated nadir (vertical) view similar to satellite imagery. The PC algorithm is implemented under the assumption that the query images and the reference images are largely overlapped. This allows the shift  $(a, b)$  to be resolved at integers level by inverse Fourier transforms (IFT). In theory, if the query image has at least a  $\frac{1}{4}$  overlap with the reference image, only one round of the PC algorithm is necessary to localize the UAV. However, in practice the overlap can be smaller and the authors implemented a scanning algorithm with a sliding window to achieve a more robust localization. This is where this approach differs from other template matching approaches such as NCC and MI. For these other approaches, scanning is not optional and is inherently required to obtain a correct localization solution. In the authors' solution, the final position of the UAV is determined by the weighted average of the candidate positions from the scanning with their normalized peak PC values. If the average of the peak PC values is larger than 0.85, the position estimation is deemed reliable. If it is not the case, the current frame is ignored and the algorithm is repeated on the following frames. In the case of a series of match failures, the search area on the reference map is dynamically enlarged. The solution was tested in multiple simulated scenarios and in one semi-simulated scenario with real UAV imagery but no inertial data. In the experiment with real UAV imagery, an unspecified length flight was performed at an altitude of 350 m leading to an image resolution of approximately 6 cm. The PC algorithm was tested against the NCC and the MI template matching localization algorithms. The results have shown an average error of 1.31 m for PC, 2.19 m for NCC and 3.08 m for MI. These results are calculated from pixels valued errors reported by the authors using the provided imagery resolution.

Patel [83] recently proposed a new localization approach building on previous work from Yol *et al.* [72] and Warren *et al.* [68]. Imagery from a quadcopter equipped with a gimballed stereo camera system was registered to ref-

erence imagery from Google Earth™ [63]. The authors based their approach on the normalized information distance (NID) [84] which is obtained from the MI similarity measure. By contrast to MI, NID is not as dependant on the amount of overlap between images making it more robust for localization applications. Similar to Yol *et al.*, the sRt warping that maximizes the NID between the reference map and the UAV image is computed by numerical optimization. For this purpose, the authors employed a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm, a quasi-Newton optimization method [85]. The optimization is performed first on Gaussian-blurred images and then refined by using the result as the initialization for an optimization pass on the raw images. This approach was shown to give similar results to a grid search with a lot fewer computations. To reduce the search space in the reference map, the previous localization procedure is applied in a radius around the predicted pose of a VO system. The initial radius is made larger and is shrinking for the following registrations. If the registration fails for multiple frames, the radius is grown to correct the drift. The VO system is based on a visual teach and repeat (VT&R) framework and is the same as the one used by Warren *et al.* [68]. The main limitation of this work is the fact that the rotational change between the UAV frame and the reference imagery is assumed to be 0. The localization solution was tested outdoor on a 1132 m flight trajectory at an altitude varying between 36 m and 42 m, with six different levels of illumination. The obtained average RMSE across all levels of illumination was 1.6 m, 0.88 m, and 1.17 m for longitude, latitude, and altitude, respectively.

### 3.2. Feature points matching

Feature points matching or indirect matching has been employed as an efficient alternative to template matching. The process involves two tasks: feature points detection and descriptor extraction. Feature points detection often involves a corner detector such as the well-known Harris [86] and FAST [87, 88] detectors. The goal of the feature points detection is to find salient points that are susceptible to be detected in two completely independent detection iterations done on different images of the same area that may have a large difference in illumination, scale, rotation and viewpoint. The descriptor extraction is the step where feature vectors are extracted from the area around the feature points. Approaches vary but notable techniques include the gradient histograms from SIFT [89] and the binary tests from BRIEF [52]. The objective is to construct a descriptor that will be used to match feature points using a metric such as the Euclidean distance or the Hamming distance. An illustration of a feature points matching between a UAV image and a satellite image is available in figure 7. A majority of the works in this section use a combination of feature points matching and statistical filtering to perform UAV localization. According to the results and considering their maturity, these approaches seems to be

the current state-of-the-art in absolute visual localization for UAVs.

The work of Seema *et al.* [90] and Saranya *et al.* [91] are very similar and aim to compare the characteristics of NCC with those of the random sample consensus (RANSAC) [92] combined with SURF [93] for the registration of a UAV image to a global reference map. For template matching with NCC, an edge detection is first performed on both the UAV and the reference map image. For every position in the reference map, NCC is applied using the UAV image as a template. The maximal response is considered as the location of the UAV image in the reference map. For the feature points approach, SURF features are extracted from both the UAV image and the reference map and then matched. RANSAC is used to remove the outliers among the matches. The position of the UAV image in the reference map is then recovered by computing a geometric transformation between the remaining feature point matches. A very limited simulation has been performed with images from Google Maps<sup>TM</sup> [94]. A reference map and a smaller piece of the map were selected arbitrarily for the experiment. Different variations of noise, scale change and rotations have been applied to the map piece to simulate an image from a UAV. The results indicated a very precise localization which is expected as the simulated UAV image is extracted from the reference map. More interestingly, the results have shown that the execution time of NCC is better than RANSAC for this experiment. However, NCC is very sensitive to scale change, rotation, noise and blur which is not the case for the RANSAC feature-based approach. The authors concluded that NCC was better suited for situations where the rotation and the scale can be obtained from other sources. Two downsides of RANSAC with SURF were identified: the requirement for a minimum number of features and the fact that the feature detection and extraction times are not constant.

Shan *et al.* [95] developed a framework that combine histograms of oriented gradients (HOG) [96], particle filtering and optical flow (OF) [97]. The reference data used in the system is constructed by applying HOG at all possible positions of a global reference map of the environment. The first step of the localization is a global localization step needed to initialize the particle filter. To avoid a costly sliding window search, a correlation filter based on 2D Fourier transforms [98] is used to obtain a global position confidence map. The location of the maximum is considered as the take off position of the UAV and the first set of particles is drawn around it. For the measurement update step, a subset of the reference data is extracted around each particle. Similarly to the reference map, HOG descriptors are also extracted at each pixel position of the UAV image. The HOG vector distances are computed between the particle subsets of reference data and the UAV image. These distances are then Gaussian normalized to form a valid probability density function (PDF). The PDF is used to compute the position estimation of the UAV us-

ing a weighted average of the particle positions. For the propagation step of the particle filter, the authors opted to use OF instead of the conventional kinematic system found in similar AVL contributions. However, the OF approach still assumes the availability of rotational motion data from an IMU and altitude estimations from a barometer which are used to determine the inter-frame translation. After propagation, a coarse to fine search is performed around the particles to reduce the computation time. The coarse search is considering a wider area around the particles and the fine search a smaller one. First, the coarse search is used and if the average HOG descriptor distance is under a threshold, the current estimation become the new estimated UAV position. If not, a fine search is performed and the same threshold is applied. Similarly, if the distance is under the threshold, the estimation is used as the current position of the UAV. If both searches fail the threshold test, the OF translation prediction is added to the previous estimation and the result is used as the current UAV position. Experiments were performed with real flight data from a UAV operating in a 40 m × 225 m environment. The authors compared their particle filter solution with a solution using only OF. An RMSE of 6.77 m was obtained with the particle filter and an RMSE of 169.19 m was obtained with the OF-only solution.

Chiu *et al.* [99] presented a complete solution for GPS-denied aerial navigation using a sensor fusion between an IMU and the absolute localization provided by image registration with geo-referenced imagery. The sensor fusion happens inside a smoother-based inference framework called sliding window factor graphs [100]. It was created from [101] and extended to deal with low-frequency vision-based measures. In this framework, the aircraft state is estimated through a fixed length sliding window designed to have a constant computation time. The framework iteratively relinearize highly nonlinear measurements of the monocular camera to obtain a better estimation of the state of the aircraft. Two different estimation processes were proposed: 2D-3D tie-points and geo-registered feature tracks. The first estimation process match an undisclosed type of features from the aerial images to a 2D reference map generated from a 3D terrain model by another subsystem. This process provides 3D absolute information to the framework. The second process expands the first one by tracking features from frame to frame allowing for the propagation of 3D absolute information. This adds a significant improvement to the accuracy of the pose estimation. Real life experiments have been performed with an airplane in two large outdoor environments. Ground truth was obtained from a real-time kinematic (RTK) differential GPS. The first flight was on a length of 38.9 km and included forest and urban areas. The second flight was 26.5 km long and over urban areas only. With the 2D-3D tie-points approach alone, the authors obtained an RMSE of 13.98 m and 10.52 m for each flight respectively. By adding the geo-registered feature tracks the RMSE was reduced to 9.83 m and 9.35 m respectively.

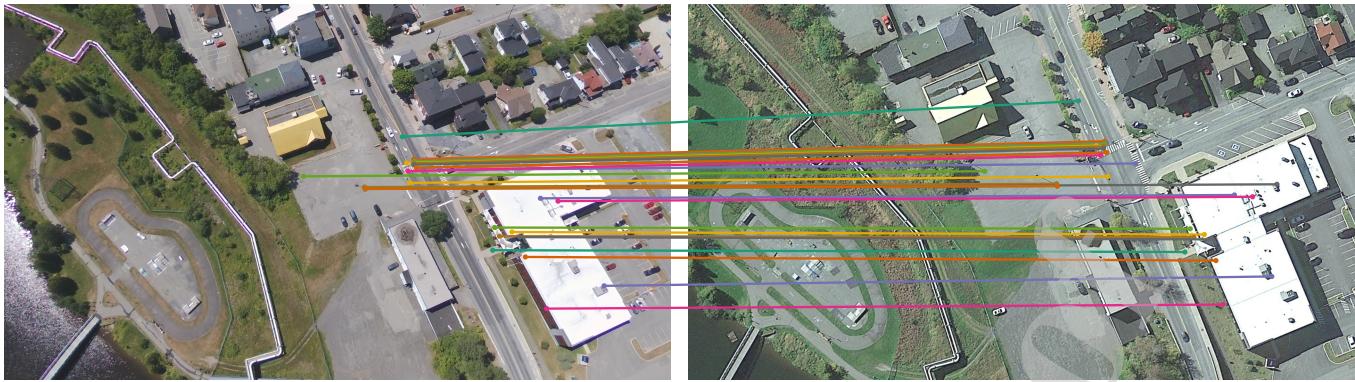


Figure 7: Example of feature points matching between a UAV image (left) and a satellite image (right).

Mantelli *et al.* [102] designed a system to perform absolute localization in 4 DoF using satellite imagery. The system is using a down-facing monocular camera with the roll and pitch angles considered close to zero. UAV images are matched against a satellite map using the authors' own version of BRIEF [52], abBRIEF. The abBRIEF descriptor is using the a and b channels from the Lab (or CIELAB) color representation instead of a grayscale image as in BRIEF. Another difference is the use of quantization levels instead of a Gaussian blur to reduce noise. This prevents pixel averaging and ensures that pixel differences are conserved. The abBRIEF descriptor is applied non-traditionally by the authors, a fixed number of pixel pairs is randomly selected over an entire image instead of being selected in the area around feature points. This produce a global descriptor for each image and allow the authors to skip corner detection to reduce the computation time. A monte carlo localization (MCL) approach, also known as particle filter, is used to estimate the location of the UAV in the state space. The measurement model is simply composed of the Hamming distances between the particle descriptors and the current UAV image descriptor. The motion update (propagation) is based on the estimations provided by a VO system. However, since the motion model is not the focus of this work, the VO solution is very basic and inaccurate. For instance, no error model was considered for this VO system. However, the authors applied a random Gaussian noise to the measurement model. The solution was tested on 3 trajectories in a large environment and with a variation of 5 reference maps. The longest trajectory was 2.4 km and the localization was performed against a map of 1.1 km × 1.1 km. Reported results indicated an average error of 17.78 m on the longest trajectory.

Masselli *et al.* [103] proposed an approach based on terrain classification and particle filtering. Patches of terrain are classified in four arbitrary classes: grass, bushes, roads and buildings. To do so, an ORB [104] feature descriptor set is extracted from each cell of a grid defined over an image. Each of these descriptors is then classified using a random forest model [105] pretrained with images

from Google Maps<sup>TM</sup> [94]. The class that is associated with the most feature descriptors is considered as the type of terrain present in the cell. This classification process is then applied to other Google Map<sup>TM</sup> [94] data to build a reference map of the target environment. The same terrain classification approach is applied to the UAV images and the localization of the UAV is estimated. For this purpose, the reference map and the processed UAV image are used inside an MCL system (particle filter). The estimated position of the UAV is computed as the mean position of 15 particles with the best weights of a set of 500 particles. This estimated position is averaged over the last four frames to reduce noise. The authors provided very little details about the particle filter implementation. However, they mentioned that IMU data was not used and that the estimations are based on vision only. Real life experiments were performed in a 60 m × 100 m outdoor environment and the average position estimation error obtained was 9.5 m.

Shan and Charan [106] developed a method that uses simultaneous feature detection and extraction to match UAV images with a reference map. Feature points are detected with maximal self-dissimilarities (MSD) [107] and the descriptors are extracted with local self-similarities (LSS) [108]. MSD is used to detect the feature points based on the rarity of the patch and LSS is used to compute a correlation surface around the feature points that is transformed in a descriptor with log-polar binning. Both algorithms use the sum of squared differences (SSD) internally as a similarity measure allowing for the recycling of the SSD values and the simultaneous execution of the detection of feature points and the extraction of their descriptors. A reference map of the environment is constructed from Google Maps<sup>TM</sup> [94] and feature points are extracted over it. OF is used to reduce the search space of the reference map to a limited area. Then, feature points are matched inside a sliding window over the area around the position predicted with OF. For each possible window, the sum of the Euclidean distances of the descriptors of the matching feature points is computed. The window with a minimal sum is considered to be the position of the UAV.

Furthermore, estimated positions that are over a certain distance from the position predicted by OF are rejected as outliers. Real flight data from a UAV was used to validate the approach. No average position estimation error was reported but a flight path comparison diagram has shown the results to be relatively close to GPS positioning and far better than OF alone.

Couturier and Akhloufi [65] proposed a new framework for RVL based on AVL. An illustration of the framework is available in figure 8. The localization task is separated in two processes: data collection and pose estimation. The data collection is accumulating sets of feature points descriptors with their corresponding geographic coordinates. This data is then used in the pose estimation process to perform localization in 2 DoF,  $(x, y)$ . The approach is used in the context of RVL to perform backtracking or return home in case of GNSS failure or GNSS signal degradation. However, even if the approach is presented as a relative approach, it can also be seen as an absolute approach with the reference map being the previously observed path. In fact, the data collection procedure can be directly replaced with satellite imagery to produce an AVL solution. The authors first developed an approach called Naive-RVL as a proof of concept of their framework. Naive-RVL was performing vision-only localization by matching the descriptors extracted from the current view of the UAV with all the descriptors of the descriptor sets present in the collected database. An accumulator associated with each descriptor set in the database was incremented each time it contained a descriptor with minimal distance from a descriptor extracted from the current UAV frame. The geographic coordinates of the descriptor set with the largest accumulator were used as an estimation of the UAV position. The authors evaluated six feature points techniques (SIFT [89], SURF [93], KAZE [109], AKAZE [110], BRISK [111] and ORB [104]) with this framework while varying the amount of extracted feature points. ORB was the method that provided the best position estimation with a mean absolute error of 68 m on a 4 km trajectory. The same authors in [66] further refined their approach by introducing a particle filter with a motion model based on the UAV's IMU and a measurement model using their vote-based feature point matching. The particle filter allowed the reduction of the database search space to 10 images on average, greatly reducing the computation time and the false positive rate. This eliminated the dependency of the computation time on the environment size and enabled a bounded execution time. The results obtained indicated an average error of 34 m on a 4 km linear trajectory with typical IMU noise. In a second refinement of their approach, the authors proposed the addition of a conditional criterion to the particle filter [67]. The criterion is being used in the measurement update step of the particle filter in order to take into consideration the feature point matches quality. The threshold used for the criterion is computed by leveraging a previously unused computation opportunity during the data collection pro-

cedure. This refinement lead to a faster, more accurate, and more stable localization solution by eliminating most of the remaining false positive feature point matches. Furthermore, the authors re-evaluated the same six feature points techniques as in their first paper. A lower mean absolute error of 24.97 m was obtained with SURF on the same 4 km trajectory.

### 3.3. Deep learning

In recent years, the field of computer vision changed drastically after a convolutional neural network (CNN) architecture named AlexNet [112] won the large scale visual recognition competition (ILSVRC) of ImageNet [113]. Researchers started to realize the potential of this new technology and started to apply CNNs to a wide variety of computer vision tasks including visual localization. However, even if a lot of areas of computer vision and robotics are now dominated by deep learning, the application to UAV localization is still in its infancy. This is mostly due to the difficulty of building an end-to-end architecture for localization. For instance, with RVL approaches making use of reference maps (e.g. [65–68]), the dynamic aspect of the input data makes it difficult to apply CNNs as they are designed to be trained offline. Most of the time CNNs are too slow for real-time training making it impossible to remember recently overflowed areas. With pure AVL approaches, the amount of data seems to make it impractical to train CNNs to remember complete reference maps accurately due to current GPUs memory limitations. The lack of public datasets for UAV localization might also be a contributing factor to the slow development of deep learning in the field. Due to these reasons, very few contributions using deep learning have been made for RVL and AVL. For this reason, we have relaxed the definition of AVL and included works that while not being exactly AVL are still worthy of mention (i.e. [114, 115], [116]). In this section, we review the available literature on the subject and draw a clear picture of the advancements in that matter.

Amer *et al.* [117] presented the concept of deep urban signatures in which a CNN is used for the computation of a unique characterization of different urban areas based on their visual appearance. The training and the validation of the CNN is performed with data from Google Maps™ [94] and then, the model is tested with data from Bing Maps™ [75] to simulate UAV imagery. The idea of this work is to exploit the differences between the structure and the organization of different urban areas or districts through a classification approach. Indeed, the urban development planning (or the lack of), the rate of development, the building materials, etc. can be significantly different from an area to another. The approach consists of two steps: district-level localization and neighborhood-level localization. A single VGG16 [118] pretrained on the ImageNet dataset was used in this localization framework. Transfer learning was used and only the fully connected layers were trained. In the district-level localization step, the network was trained to classify the query images in one of

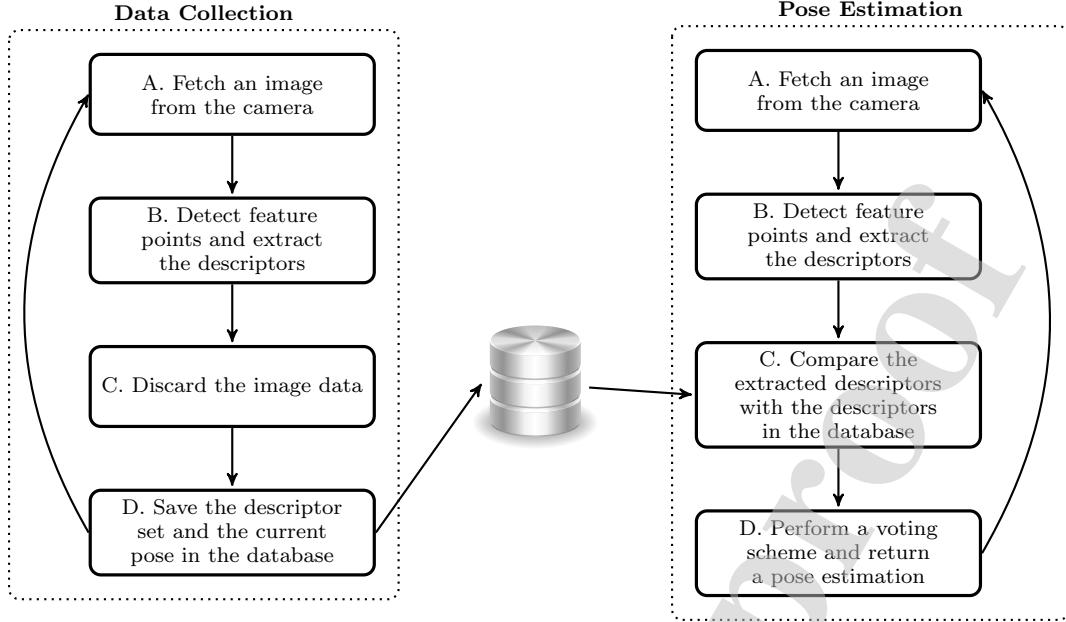


Figure 8: High-level overview of the RVL framework.

the seven districts in order to reduce the search space. For the neighborhood-level localization step, a nearest neighbor search is performed using feature maps from the same CNN. For this purpose, the feature maps from the 14<sup>th</sup> convolutional layer were extracted for the query image and the reference images and compared. The position of the reference image closest to the query image was used as the UAV position estimation. Experiments has been performed with the images from seven different districts in Cairo, Egypt. An average district classification accuracy of 91.2% and an average position estimation error of 200.75 m were obtained. The main drawback of this approach is the need to train the network in a specific area before using the localization solution.

Nassar *et al.* [114, 115] combined traditional computer vision techniques with CNNs in a single framework. A known initial position is used to define a window,  $r$ , on a reference satellite map,  $R$ . The first step is a calibration which is performed every three frames to keep the current view of the UAV registered with the reference map. For this purpose, SIFT [89] feature points are detected and their descriptor extracted from both the UAV image and  $r$ . Then, RANSAC [92] is used to estimate the homography matrix that describes the transformation from  $r$  to the UAV image. The computed homography is applied to  $r$  and the resulting image is passed to the next step called sequential frame registration. Sequential frame registration is executed at processing rates, which means as often as possible. Similarly to the calibration step, the current UAV image is registered with the calibrated  $r$  using ORB [104]. The resulting homography is used to update  $r$ . The use of SIFT every three frames and ORB at every frame allow the registration to be both computationally efficient

and accurate. Indeed, SIFT is better at handling scale variance and ORB is a lot faster than SIFT. Thus, the authors' approach can be seen as a simple way to combine the two feature extraction methods. In the next step of the framework, semantic segmentation is performed using U-Net [119]. The UAV image and the registered  $r$  are fed to U-Net to obtain precise masks of buildings and roads. The masks are refined with the morphological operations open and close in order to fill gaps in blobs and remove noise. Small shapes and blobs are filtered out and Hu moments [120] are computed on the remaining blobs. The blobs are then brute force matched between both images using the computed moments. A scoring system is used to keep only the best matches which are used to compute a refined homography. The approach was evaluated against two datasets captured along paths of 1.2 km and 0.5 km from two different cities. The average geolocation error (longitude-latitude) obtained with local features matching is 10.4 m and 6.3 m for each dataset respectively. The addition of the deep learning based semantic segmentation reduce the average error to 5.1 m and 3.6 m respectively. However, this addition limits the usage of the localization system to known urban areas. Furthermore, this approach is not exactly an AVL solution as the UAV image is kept registered with the reference map at all times. This limits the exploration of the possible positions and induce drift in the solution.

Marcu *et al.* [121] proposed a multi-stage multi-task (MSMT) neural network architecture to simultaneously perform aerial image segmentation and geolocalization in a single pass. The network is using a single encoder and two decoders to fulfill its tasks. The first stage is a semantic segmentation network based on [122] that is used

to obtain road segmentation. According to the authors, it can act as a unique footprint of an urban area. As illustrated in figure 9, the remaining of the network is divided in two branches. The first branch (i.e. LocDecoder-R-2) is learning to identify the location of the input image and to output a corresponding longitude-latitude pair. In the second branch (i.e. LocDecoder-S-128), another segmentation is generated in which white pixels represent the possible locations of the UAV. The position of the UAV is estimated by first considering the centroid of the largest connected component in the segmentation mask. Then, if the segmentation fails, the output from the longitude-latitude regressor is used. The architecture is trained in multiple steps. First the UniEncoder-512 encoder and the SegDecoder-512 decoder are trained together until convergence and then their weights are frozen. Following this, the decoders of the two parallel branches are trained independently using the UniEncoder-512 as their encoder. The longitude-latitude regressor is trained using an RMSE loss. The loss function used for the segmentation pathways is a linear combination of the binary cross-entropy and the dice loss functions. Separate from the network, an extra alignment refinement step using the iterative closest point (ICP) algorithm [123] is added. Its purpose is to precisely align the road segmentation with the OpenStreetMap™ [124] roads near the estimated UAV position. Experiments were realized in the form of a simulation by using random  $100\text{ m} \times 100\text{ m}$  patches around intersections in a  $70\text{ km}^2$  European urban area. The constructed dataset was split in 90% for training and 10% for testing. Without alignment, 96.84% of test locations had an error of less than 20 m. With alignment 94.56% of the test locations are within 2.5 m of the ground truth location and 97.58% are within 5 meters. This approach has two major drawbacks. The first is the use of road segmentation that mostly limit the solution to urban areas. The second is the fact that the longitude-latitude regressor must be trained with images from the area of the flight prior to the flight.

Goforth and Lucey [116] proposed an approach to AVL combining a CNN and VO inside an optimization framework. The approach is inspired by [125, 126] and allows to learn a deep feature representation for direct image alignment. The network is composed of two parallel CNNs built from the first 3 layers of a VGG16 [118] and containing the same weights. Two images to be aligned are fed into the network, one in each CNN. Resulting feature images are then passed to a differentiable implementation of the Inverse Compositional Lucas-Kanade (ICLK) algorithm [127] that allows for backpropagation. The ICLK layer is implemented as in [127] using full projective alignment with homography and a geometric corner loss function. The architecture is trained using only freely available satellite imagery from the United States Geological Survey Earth Explorer website [128]. It contains ten large images containing a mix of urban, suburban, and rural areas captured between 2006 and 2017 across different seasons. The training data is generated by randomly se-

lecting corresponding patches in two of the images. The first patch is kept static and the second one is warped using random projective wrap coordinates. Resulting data is divided in 80% for training and 20% for testing. The location of the UAV is estimated by composing a series of homography computed along the way with the CNN. This aspect makes this approach closely related to VO which means that it is highly susceptible to drift over time. To help alleviate this issue the authors added a pose parameter optimization over a sliding window. This addition is inspired by SLAM and photometric bundle adjustment [129, 130]. However, it only helps reduce the drift and does not eliminate it. Therefore, this approach could be qualified as a VO technique with satellite imagery prior that would stand somewhere between RVL and AVL. The solution was validated against two freely available UAV sequences captured by senseFly [131]. The first flight is at an altitude of 200 m and over a path of 850 m, the second flight is at an altitude of 220 m and over a path of 610 m. Reference data for this validation is extracted from Google Earth™ [63]. Results indicate an average error in x-y of 7.06 m and 25 m and an average error in altitude of 8.01 m and 7.70 m for the first and second datasets respectively.

Schleiss [132] developed an approach using conditional generative adversarial network (cGAN) [133, 134] and template matching. In contrast with some other works, the author intentionally avoided embedding any concepts of VO in its solution in order to obtain a pure AVL approach. The first step of the proposed algorithm is to transform the current image from the UAV to a map-like representation similar to geographic maps without satellite imagery overlay. This is achieved using a cGAN trained to segment images in three classes: buildings, roads, and background. For the training, satellite data is extracted from OpenStreetMap™ [124] and corresponding metadata is used to generate matching map-like representations. The training dataset is composed of  $125\text{ km}^2$  of map data and the testing dataset is composed of  $9\text{ km}^2$  of map data. Once generated, the map-like image is matched against a similar map-like reference map of the region using a template matching technique. The template matching technique is using the normalized sum of squared differences (SSD) and is inspired by [135]. It is assumed that the scale and orientation of the template (map-like UAV image) are known and that they can be obtained with an IMU. Therefore, the template is first scaled and rotated to match the reference map. Then, SSD values are computed between the template for each possible position in the reference map. The position with the lowest value is considered as the estimation of the UAV position. For the evaluation of the solution, the reference map was collected by the author with a plane flying in a grid trajectory over an area of  $560\text{ m} \times 680\text{ m}$ . Ground truth was obtained using an RTK GPS. The same images from the reference map were used to construct a simulated 1.61 km long UAV flight. By removing the images without enough features from the reference map, an average error of 22.7 m in x-y can be

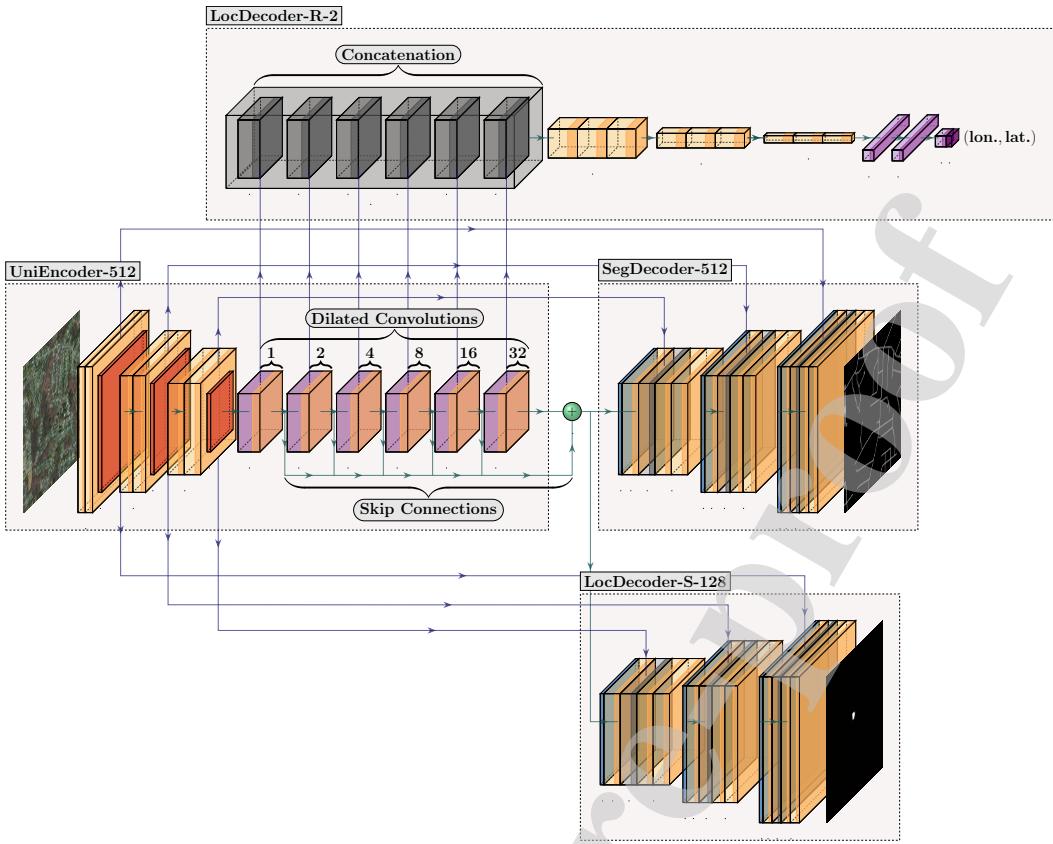


Figure 9: Multi-stage multi-task (MSMT) architecture for semantic segmentation and geolocalization.

obtained. However, when the entire map is considered, the average error is 40 m.

### 3.4. Visual Odometry

As stated in section 2, VO is an approach based on the comparison of consecutive frames captured by a camera. Differences in motion between frames are accumulated during the flight to obtain a localization estimation making it well suited for RVL. However, VO is not limited to RVL approaches and can be extended to build AVL systems by using data from previous flights. In such approaches, the data collected with VO is georeferenced and collected in a database to later be reused for localization. This approach is very different from the approaches reviewed until now as, with the noted exception of the work of Goforth and Lucey [116], none of them relies on consecutive frames comparisons. This characteristic of the approach incurs fundamental differences in how the comparison of the current UAV image with the collected database is performed during visual localization. As in these methods UAV images must often be compared to a pose graph instead of only images or image features, the localization algorithm is facing a completely different set of challenges. This motivated the creation of this fourth category of AVL approaches. To our knowledge, only two works have been

conducted in this area. The work of Goforth and Lucey [116] which is closer to deep learning and, therefore, presented in section 3.3, and the work of Warren *et al.* [68]. The originality and the uniqueness of the latter formulation was also a key factor that motivated the creation of this category. By opposition to other approaches, this approach has a lot of potential for applications that require very accurate knowledge of a single location at very low altitudes.

Warren *et al.* [68] developed an approach to allow a UAV to return to its starting position after a GPS failure. The system is using a gimbal-stabilized grayscale calibrated stereo vision system to acquire data. The authors adapted a VT&R path-following algorithm from previous unmanned ground vehicle (UGV) works. VO and GPS navigation is first used to construct a relative pose map when the UAV have a stable GPS connection. The VO system is using SURF [93] feature points which are triangulated from both the stereo pair and the motion. This allows for localization in proximity to the ground with the stereo pair and in large depth situation with motion. The feature points are matched with the help of an onboard GPU, then outliers are removed and the transform relative to the last keyframe is obtained using maximum likelihood estimation sample consensus (MLESAC) [136]. The

transform is optimized using the authors' own simultaneous trajectory estimation and mapping (STEAM) bundle adjustment engine [137]. Then, if the keyframe presents enough change in comparison to the last one, it is added to the pose graph. Windowed bundle adjustment is performed on the last 5 to 10 vertices in the graph to form a dead-reckoned path. During the return phase (without GPS), visual matching is performed with respect to the local 3D map in parallel with VO. Previous dead-reckoned paths are used to reduce the number of potential matches in this phase. Matches are treated as in VO with MLESAC and STEAM to obtain the final transform relative to the pose graph. The estimated localization is then fed to a path planning system to stay on track to the starting point. The localization has been tested in a 140 m × 160 m outdoor environment. On average, the positional error (y-z) from the original path is 1.5 m and can reach a maximum of 3.6 m in the turns.

#### 4. Results and discussion

The results of the reviewed works are difficult to compare. When it's possible to compare them, it's almost impossible to derive any meaningful interpretation of the comparison. Multiple factors are in play in this issue.

The first, and perhaps the most important issue, is the lack of a common method or standard to report results. This seems to be fairly common in new areas of research. In such a situation, the authors of new contributions are not sufficiently aware or don't have access to enough previous works. It causes a lack of information that in turn makes the authors choose different result reporting approaches. For instance, some authors report their results with the RMSE metric while others are reporting them with the average Euclidean distance (AED). While most authors are using the AED to report their results, some differences persists in the reporting of this metric. Sometimes the AED is reported for the complete test series and sometimes it is only reported after eliminating the worst results of the test series. For example, Marcu *et al.* [121] are only reporting the percentages of the tests samples under an error threshold (e.g. 96.84% have an error inferior to 20 m). This approach while being valuable for the evaluation of the average performance of the system may make the results more difficult to compare with other works. We recommend that the authors using this approach provide these results alongside the average of all test results as it has been done in [132]. Another issue with result reporting is the state space. Some authors are reporting the AED or the RMSE on x, y and z independently and others combine them and report the metric for x-y or x-y-z. When possible, we suggest reporting both the AED and the RMSE for the following combinations: x, y, z, x-y, x-y-z,  $\phi$ ,  $\theta$ , and  $\psi$ . If the error is reported using longitude, latitude and altitude, we recommend converting these coordinates to a Cartesian space and in addition to report the x, y and z errors as previously stated.

The second issue is the lack of standard benchmark datasets. Not having standard benchmark datasets force researchers to build their own custom datasets. The result is that variables such as the size of the reference map, the resolution of the reference map, the size of the test environment, the length of the flight, the altitude of the flight, the velocity of the UAV, and the sensors are different for each of the reviewed works. In such a context, even with matching metrics, it is still impossible to perform a fair comparison of two approaches. This not only slows advancement in research as researchers must gather their data themselves, it also prevents researchers from evaluating the performance of their work in comparison with the state-of-the-art. In fact, it makes it difficult to identify the state-of-the-art and it can make it difficult for the peer reviewers to evaluate the quality of a proposed work. Nevertheless, we have found attempts to standardize datasets in [121] and [116], and an attempt to perform a fair comparison with other approaches in [102]. This type of contributions are strongly recommended but to this day, they remains very rare in the field.

The third issue is the lack of public release of source code. Of all reviewed works, only the work of Goforth and Lucey [116] was released alongside its source code. As stated in section 1, the main challenges of vision-based approaches to localization reside in the vast amount of data to analyze and into the complexity of interpreting this data. These challenges implies a need for an order of magnitude more computations than traditional GNSS-INS approaches. This comes from the fact that millions of pixels intensities must be considered during computations by contrast with a few geographical coordinates in GNSS-INS approaches. In fact, this seems to be the main remaining limitation of AVL approaches. Understandably, this leads computational efficiency to be the one of the most important aspect motivating research in this field, second only to the desire of developing interference-immune localization solutions. As a matter of fact, more than half of the reviewed works have given some thoughts to computational efficiency. As it is an important aspect of the research in the field, it becomes increasingly important to compare contributions on that basis. In real-world applications, the benefits of solving GNSS-INS weaknesses must be carefully weighed against the incurred increase in computational complexity. For example, there might be more disadvantages in integrating vision-based localization in micro aerial vehicles (MAV) than benefits as their energy resources are very limited. However, in high altitude long endurance (HALE) UAVs this extra computation is mostly negligible and vision-based localization can be very beneficial. A similar example could also be constructed by comparing UAVs operating in urban areas versus UAVs that don't operate in such environments. Therefore, we believe researchers can better contribute to the field by reporting computation times, discussing the hardware platform on which these tests were conducted, and publicly releasing their source code.

The fourth issue is the variations in solution validation procedures. The most common evaluation approach among the reviewed works is offline simulations with real flight data. Nevertheless, some of the reviewed works had to take a different direction and simulate the view of the UAV with satellite imagery. This approach is necessary if no real flight data is available and can be valuable for the initial validation of a localization solution. However, the issue with this approach is that satellite data is orthorectified to obtain a nadir view that is very different from real UAV images. This difference is due to perspective changes and lens distortions present in UAV images. It can greatly influence the perceived performance of a localization system, especially if the reference map and the simulated UAV images are from the same source. In [90] and [91] this approach was used and even if noise was added, it seems to have greatly influenced the results. Furthermore, the camera sensors on satellites are very different from the CMOS-based camera sensors found on most consumer-grade UAVs. Matching imagery from different sources is at the core of AVL and we recommend that this aspect be considered in the design of validation frameworks for AVL. Note that this issue, could be alleviated with a broader availability of UAV localization datasets. Another notable validation approach we encountered, was real test flights with fully integrated localization in a closed loop system as in [68]. While being a different validation approach, it remains compatible with offline simulations using real flight data and can be beneficial to the field as it is even closer to real life applications.

As the reader might have noticed, the previously discussed issues are all in one way or another connected to results reproducibility. Differences in validation datasets, validation approaches and result reporting approaches are all factors that could, to an extent, impede on reproducibility. In addition, we note that only Goforth and Lucey released both the source code and the dataset used for their work. They did so after noticing what seemed to be issues with reproducibility and as an effort to encourage more methods and baselines in this space [116]. Following these observations, we believe that the field is being impacted by the reproducibility crisis in science [138]. However, the recommendations formulated until now can help mitigate this issue. Furthermore, we note that the release of source code and datasets is on its way to become part of standard open science practices [139].

As previously stated, the results of the reviewed works are difficult to compare between each other. Nevertheless, we still provide 4 tables as a best effort to compare the reviewed contributions. These tables are meant as a way to provide, in a concentrated form, the more variables as possible to allow for the best comparison possible. By this approach, we want to advise the reader that even if the results are reported quantitatively, the comparison of methods should consider multiple variables. Furthermore, even if multiple variables are taken into account, comparisons can be at best considered as very rough. Table 1 is

provided in order to compare the techniques used by the authors in the engineering of the localization solutions. It can be observed that feature points matching is the most used approach for AVL. Table 2 allows for a comparison of the reference imagery used in each contribution. As we can see, Google satellite imagery products are the most popular for reference maps. Table 3 compares the operating environment in which the solutions were tested. The environment size is the minimal area enclosing the flight trajectory and should not be mistaken for the reference map size. As we can see, there is a great deal of variability in the validation environments. Table 4 is an important table and contains the performance evaluation of the methods. When possible we reported only results derived from real UAV imagery, however, in some cases this type of evaluation was not performed. As we can see, the AED is the most employed metric and the two-dimensional state spaces are the most used as they encompass the majority of the challenge of AVL.

## 5. Possible research directions

Following the review of these works, multiple possible research directions are provided to guide research in the field:

- *Development of specialized feature point methods for aerial imagery.* Some authors have designed their own feature point methods and others evaluated numerous existing approaches before choosing the best approach empirically. However, research specifically aimed at developing specialized feature points detection and extraction approaches for aerial imagery is lacking and can be further explored.
- *Visual odometry AVL.* To our knowledge, only a couple works has tried to adapt VO to AVL (i.e. [116], [68]). These works have demonstrated the potential of this approach however, much work remain to be done in this area.
- *Deep learning AVL.* As mentioned before, the development of deep learning approaches for AVL is limited and severely obstructed by challenges specific to AVL and localization in general. More work in this area is needed to overcome these challenges and reduce the limitations of the current approaches.
- *Publicly available datasets for AVL.* Datasets needs to be developed to benchmark new works against existing works in the literature. This is a hard requirement for the advancement of the field. Such datasets should include satellite imagery, real UAV flight data and inertial data.
- *Freely available implementations.* Researchers in the field should try to publish source code to allow better

Table 1: Comparison of the approaches

Work	Matching type	Matching technique	Search space reduction
Dalen <i>et al.</i> [71]	Template	NCC [73] + SLAM [76]	Particle filter
Yol <i>et al.</i> [72]	Template	MI [78, 79]	Newton's optimization
Wan <i>et al.</i> [80]	Template	PC [81]	Adaptable sliding window
Patel [83]	Template	NID [84]	VT&R VO
Seema <i>et al.</i> [90] and Saranya <i>et al.</i> [91]	Feature points	SURF [93] + RANSAC [92]	None
Shan <i>et al.</i> [95]	Feature points	HOG [96]	Particle filter
Chiu <i>et al.</i> [99]	Feature points	2D-3D tie-points + feature tracks [99]	Sliding window factor graphs
Mantelli <i>et al.</i> [102]	Feature points	abBREIF [102]	Particle filter
Masselli <i>et al.</i> [103]	Superpixels	ORB [104] + random forest [105]	Particle filter
Shan and Charan [106]	Feature points	MSD [107] + LSS [108]	Optical flow
Couturier and Akhloufi [65] Naive-RVL	Feature points	ORB [104] (best)	None
Couturier and Akhloufi [66] PF-RVL	Feature points	ORB [104]	Particle filter
Couturier and Akhloufi [67] CP-RVL	Feature points	SURF [93] (best)	Particle filter
Amer <i>et al.</i> [117]	Deep learning	VGG16 [118] + Deep urban signatures [117]	District-level localization
Nassar <i>et al.</i> [114, 115] /wo seg.	Feature points	SIFT [89] + ORB [104] + RANSAC [92]	Constant registration
Nassar <i>et al.</i> [114, 115] /w seg.	Deep learning	SIFT [89] + ORB [104] + RANSAC [92] + U-Net [119] + Hu moments [120]	Constant registration
Marcu <i>et al.</i> [121] /wo alignment	Deep learning	MSMT [121]	Segmentation + lon.-lat. regressor
Marcu <i>et al.</i> [121] /w alignment	Deep learning	MSMT [121] + ICP [123]	Segmentation + lon.-lat. regressor
Goforth and Lucey [116]	Deep learning	VGG16 [118] + Deep-LK [125] + Photometric bundle adjustment [130]	VO + sliding window
Schleiss [132]	Deep learning	cGAN [133, 134] + SSD [135]	None
Warren <i>et al.</i> [68]	Feature points	SURF [93] + MLESAC [136] + STEAM BA [137]	Dead-reckoned paths

Table 2: Comparison of the reference imagery

Work	Reference imagery	Map size (m)
Dalen <i>et al.</i> [71]	Bing Maps™ [75]	-
Yol <i>et al.</i> [72]	Google Earth™ [63]	130 × 100
Wan <i>et al.</i> [80]	Ziyuan-3 satellite	-
Patel [83]	Google Earth™ [63]	262 × 262
Seema <i>et al.</i> [90] and Saranya <i>et al.</i> [91]	Google Maps™ [94]	761.6 × 761.6
Shan <i>et al.</i> [95]	Google Maps™ [94]	300 × 150
Chiu <i>et al.</i> [99]	N/D GIS (likely DARPA)	-
Mantelli <i>et al.</i> [102]	Google Earth™ [63]	1160 × 1160
Masselli <i>et al.</i> [103]	Google Maps™ [94]	150 × 90
Shan and Charan [106]	Google Maps™ [94]	-
Couturier and Akhloufi [65] Naive-RVL	Own	4000 × 243
Couturier and Akhloufi [66] PF-RVL	Own	4000 × 243
Couturier and Akhloufi [66] CP-RVL	Own	4000 × 243
Amer <i>et al.</i> [117]	Google Maps™ [94]	5251 × 9421, 3960 × 4874, 3960 × 2741, 1945 × 1523, 2132 × 2132, 3354 × 2437, and 4874 × 5178 combined
Nassar <i>et al.</i> [114, 115]	Google Earth™ [63] and Bing Maps™ [75]	2160 × 2160, 1240 × 1240, and 1130 × 1130 combined
Marcu <i>et al.</i> [121]	OpenStreetMap™ [124]	70000 × 70000
Goforth and Lucey [116]	Google Earth™ [63]	900 × 700 and 1200 × 600
Schleiss [132]	Own	560 × 680
Warren <i>et al.</i> [68]	Own	140 × 160

Table 3: Comparison of the validation environments

Work	Trajectory (m)	Environment size (m)	Altitude (m)
Dalen <i>et al.</i> [71]	525	90 × 100	61
Yol <i>et al.</i> [72]	695	-	125 to 175
Wan <i>et al.</i> [80]	-	-	350
Patel [83]	1132	250 × 250	36 to 42
Seema <i>et al.</i> [90] and Saranya <i>et al.</i> [91]	Static position	192 × 192	-
Shan <i>et al.</i> [95]	520	40 × 225	80
Chiu <i>et al.</i> [99]	38900 and 26500	-	-
Mantelli <i>et al.</i> [102]	2400	-	40 to 300
Masselli <i>et al.</i> [103]	258	60 × 100	30 to 50
Shan and Charan [106]	-	-	-
Couturier and Akhloufi [65] Naive-RVL	4000	Straight line (4000 × 243 from FOV)	150
Couturier and Akhloufi [66] PF-RVL	4000	Straight line (4000 × 243 from FOV)	150
Couturier and Akhloufi [66] CP-RVL	4000	Straight line (4000 × 243 from FOV)	150
Amer <i>et al.</i> [117]	Multiple static positions	N × 500 × 500	-
Nassar <i>et al.</i> [114, 115]	1200 and 500	-	300 and -
Marcu <i>et al.</i> [121]	Multiple static positions	N × 100 × 100	-
Goforth and Lucey [116]	850 and 610	Straight lines	200 and 220
Schleiss [132]	1610	-	300
Warren <i>et al.</i> [68]	446	140 × 160	12

Table 4: Comparison of the performance

Work	Type of experiment	Metric	State space	Average error (m)
Dalen <i>et al.</i> [71]	Real data	Euclidean	(lon., lat.)	3.6
Yol <i>et al.</i> [72]	Real data	RMSE	(lon., lat., alt., $\phi$ )	8.02 (lon.), 6.56 (lat.), 7.44 (alt.)
Wan <i>et al.</i> [80]	Real data	Euclidean	( $x, y$ )	1.31
Patel [83]	Real data	RMSE	(lon., lat., alt., $\phi$ )	1.6 (lon.), 0.88 (lat.), 1.17 (alt.)
Seema <i>et al.</i> [90] and Saranya <i>et al.</i> [91]	Limited simulation	Absolute	(lon., lat.)	0.25 (lon.), 0.11 (lat.)
Shan <i>et al.</i> [95]	Real data	RMSE	( $x, y$ )	6.77
Chiu <i>et al.</i> [99]	Real data	RMSE	( $x, y, z$ )	9.83 and 9.35
Mantelli <i>et al.</i> [102]	Real data	Euclidean	(lon., lat., alt., $\phi$ )	17.78 (lon., lat., alt.)
Masselli <i>et al.</i> [103]	Real data	Euclidean	( $x, y, z, \phi$ )	9.5 ( $x, y$ )
Shan and Charan [106]	Real data	Qualitative	( $x, y$ )	Close to GPS
Couturier and Akhloufi [65] Naive-RVL	Real data	Euclidean	( $x, y$ )	68
Couturier and Akhloufi [66] PF-RVL	Real data	Euclidean	( $x, y$ )	34
Couturier and Akhloufi [66] CP-RVL	Real data	Euclidean	( $x, y$ )	24.97
Amer <i>et al.</i> [117]	Simulation	Euclidean	(lon., lat.)	200.75
Nassar <i>et al.</i> [114, 115] /wo seg.	Simulation and real data	Euclidean	(lon., lat.)	10.4 and 6.3
Nassar <i>et al.</i> [114, 115] /w seg.	Simulation and real data	Euclidean	(lon., lat.)	5.1 and 3.6
Marcu <i>et al.</i> [121] /wo alignment	Simulation	Euclidean	(lon., lat.)	96.84% in 20
Marcu <i>et al.</i> [121] /w alignment	Simulation	Euclidean	(lon., lat.)	97.58% in 5
Goforth and Lucey [116]	Real data	Euclidean	( $x, y, z, \phi, \theta, \psi$ )	7.06 and 25 ( $x, y$ ), 8.01 and 7.70 ( $z$ )
Schleiss [132]	Real data	Euclidean	(lon., lat.)	40
Warren <i>et al.</i> [68]	Real data	Euclidean	( $x, y, z, \phi, \theta, \psi$ )	1.5 (max at 3.6) ( $x, y$ )

reproducibility. As this practice is almost nonexistent in the field, it is expected to make future contributions a lot more valuable.

- *Embedded GPU implementations.* Very few authors in the field explored embedded and real-time localization solutions. Considering recent advances in affordable and lightweight GPUs [140], this will become an increasingly interesting area of research.
- *Hyper-spectral AVL.* Very few works have been done in the area of hyper-spectral localization and to our knowledge none in the field of AVL. For example, registering short-wave infrared (SWIR) imagery to RGB satellite reference maps could be of great interest as it would allow for vision-based localization in the presence of clouds and fog.

## 6. Conclusion

A review of the literature pertaining to absolute visual localization for UAVs was realized. The problematic was detailed with examples and figures illustrating the drawbacks of GNSS-based localization. Approaches for AVL in the literature have been deeply analyzed and synthesized to convey relevant information to allow for easy identification of contributions and to provide insight into the current state of the field. The literature was divided in four categories to allow for a better organization and visualization of the contributions in the field. Results and issues obstructing productive work in the field were assessed and discussed in detail and in a constructive manner. Four tables were presented with the objective of providing the best possible means of comparison between the works of the literature. Even if direct comparison of works is difficult due to the variance in validation approaches and result reporting techniques, this represents a best effort toward this objective. Furthermore, seven research directions were identified as a result of this review. These directions will allow the readers to better grasps the challenges of the field and help researchers to identify possible topics for their next works. On a final note, even considering that the work in the field has started more than a decade ago, the field remains in its infancy. Much work remains to be done, and in our opinion the only way to achieve it efficiently and productively will be through the standardization of solution validation and through publicly available source code and datasets.

## References

- [1] Drones: Reporting for Work, <http://www.goldmansachs.com/insights/technology-driving-innovation/drones/> (2016).
- [2] N. Poponak, Drones: Flying into the Mainstream, <https://www.goldmansachs.com/insights/pages/drones-flying-into-the-mainstream.html> (May 2016).
- [3] Commercial drones are the fastest-growing part of the market, <https://www.economist.com/technology-quarterly/2017/06/08/commercial-drones-are-the-fastest-growing-part-of-the-market> (Jun. 2017).
- [4] D. Gettinger, Summary of Drone Spending in the FY 2019 Defense Budget Request, Center for the Study of the Drone at Bard College (Apr. 2018).
- [5] J. F. Keane, S. S. Carr, A brief history of early unmanned aircraft, Johns Hopkins APL Technical Digest 32 (3) (2013) 558–571.
- [6] D. Bein, W. Bein, A. Karki, B. B. Madan, Optimizing Border Patrol Operations Using Unmanned Aerial Vehicles, in: 2015 12th International Conference on Information Technology - New Generations, 2015, pp. 479–484. doi:10.1109/ITNG.2015.83.
- [7] J. Scherer, S. Yahyanejad, S. Hayat, E. Yanmaz, T. Andre, A. Khan, V. Vukadinovic, C. Bettstetter, H. Hellwagner, B. Rinner, An Autonomous Multi-UAV System for Search and Rescue, in: Proceedings of the First Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use, DroNet '15, ACM, New York, NY, USA, 2015, pp. 33–38. doi:10.1145/2750675.2750683.
- [8] J. C. Hodgson, S. M. Baylis, R. Mott, A. Herrod, R. H. Clarke, Precision wildlife monitoring using unmanned aerial vehicles, Scientific Reports 6 (2016) 22574. doi:10.1038/srep22574.
- [9] L. F. Gonzalez, G. A. Montes, E. Puig, S. Johnson, K. Mengersen, K. J. Gaston, Unmanned Aerial Vehicles (UAVs) and Artificial Intelligence Revolutionizing Wildlife Monitoring and Conservation, Sensors 16 (1) (2016) 97. doi:10.3390/s16010097.
- [10] M. A. Akhloufi, N. A. Castro, A. Couturier, UAVs for wildland fires, in: Autonomous Systems: Sensors, Vehicles, Security, and the Internet of Everything, Vol. 10643, International Society for Optics and Photonics, 2018, p. 106430M. doi:10.1117/12.2304834.
- [11] M. A. Akhloufi, N. A. Castro, A. Couturier, Unmanned aerial systems for wildland and forest fires: Sensing, perception, cooperation and assistance, arXiv:2004.13883 [cs]doi:10.1117/12.2304834.
- [12] S. Candiago, F. Remondino, M. De Giglio, M. Dubbini, M. Gattelli, Evaluating Multispectral Images and Vegetation Indices for Precision Farming Applications from UAV Images, Remote Sensing 7 (4) (2015) 4026–4047. doi:10.3390/rs70404026.
- [13] S. Jordan, J. Moore, S. Hovet, J. Box, J. Perry, K. Kirsche, D. Lewis, Z. T. H. Tse, State-of-the-art technologies for UAV inspections, IET Radar, Sonar & Navigation 12 (2) (2017) 151–164. doi:10.1049/iet-rsn.2017.0251.
- [14] I. L. Turner, M. D. Harley, C. D. Drummond, UAVs for coastal surveying, Coastal Engineering 114 (2016) 19–24. doi:10.1016/j.coastaleng.2016.03.011.
- [15] R. H. Fraser, I. Olthof, T. C. Lantz, C. Schmitt, UAV photogrammetry for mapping vegetation in the low-Arctic, Arctic Science 2 (3) (2016) 79–102. doi:10.1139/as-2016-0008.
- [16] J. A. Goncalves, R. Henriques, UAV photogrammetry for topographic monitoring of coastal areas, ISPRS Journal of Photogrammetry and Remote Sensing 104 (2015) 101–111. doi:10.1016/j.isprsjprs.2015.02.009.
- [17] M. Jun, S. I. Roumeliotis, G. S. Sukhatme, State estimation of an autonomous helicopter using kalman filtering, in: Proceedings 1999 IEEE/RSJ International Conference on Intelligent Robots and Systems. Human and Environment Friendly Robots with High Intelligence and Emotional Quotients (Cat. No.99CH36289), Vol. 3, 1999, pp. 1346–1353. doi:10.1109/IROS.1999.811667.
- [18] Y. Oshman, I. Shaviv, Optimal tuning of a Kalman filter using genetic algorithms, in: AIAA Guidance, Navigation, and Control Conference and Exhibit, 2000, p. 4558. doi:10.2514/6.2000-4558.
- [19] J. Sasiadek, Q. Wang, R. Johnson, L. Sun, J. Zalewski, Uav navigation based on parallel extended Kalman filter, in: AIAA Guidance, Navigation, and Control Conference and Exhibit, 2000, p. 4165. doi:10.2514/6.2000-4165.
- [20] InvenSense Inc., MPU-6000 and MPU-6050 Product

- Specification, 3rd Edition (2013).
- [21] Parrot bebop 2: The lightweight, compact hd video drone, <https://www.parrot.com/us/drones/parrot-bebop-2> (2017).
- [22] B. Hofmann-Wellenhof, H. Lichtenegger, E. Wasle, GNSS Global Navigation Satellite Systems: GPS, GLONASS, Galileo, and more, Springer-Verlag, Wien, 2008.
- [23] H. Kuusniemi, G. Lachapelle, GNSS signal reliability testing in urban and indoor environments, in: Proceedings of the NTM Conference, 2004.
- [24] G. Lachapelle, GNSS Indoor Location Technologies, *Journal of Global Positioning Systems* 3 (1&2) (2004) 2–11. doi:[10.5081/jgps.3.1.2](https://doi.org/10.5081/jgps.3.1.2).
- [25] I. Ballester-Grpide, E. Herriz-Monseco, A. Miguel, M. Romay-Merino, T. Beech, Future GNSS constellation performances inside urban environments, in: Proceedings of the 13th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 2000), 2000, pp. 2436–2445.
- [26] G. Lachapelle, S. Ryan, M. Petovello, J. Stephen, Augmentation of GPS/GLONASS for vehicular navigation under signal masking, in: Proceedings of the 10th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GPS 1997), 1997, pp. 1511–1519.
- [27] M. Tsakiri, M. Stewart, T. Forward, D. Sandison, J. Walker, Urban Fleet Monitoring with GPS and GLONASS, *The Journal of Navigation* 51 (3) (1998) 382–393. doi:[10.1017/S0373463398007929](https://doi.org/10.1017/S0373463398007929).
- [28] S. Ji, W. Chen, X. Ding, Y. Chen, C. Zhao, C. Hu, Potential Benefits of GPS/GLONASS/GALILEO Integration in an Urban Canyon Hong Kong, *The Journal of Navigation* 63 (4) (2010) 681–693. doi:[10.1017/S0373463310000081](https://doi.org/10.1017/S0373463310000081).
- [29] N. Zhu, J. Marais, D. Btaille, M. Berbineau, GNSS Position Integrity in Urban Environments: A Review of Literature, *IEEE Transactions on Intelligent Transportation Systems* PP (99) (2018) 1–17. doi:[10.1109/TITS.2017.2766768](https://doi.org/10.1109/TITS.2017.2766768).
- [30] L. Wang, P. D. Groves, M. K. Ziebart, Multi-Constellation GNSS Performance Evaluation for Urban Canyons Using Large Virtual Reality City Models, *The Journal of Navigation* 65 (3) (2012) 459–476. doi:[10.1017/S0373463312000082](https://doi.org/10.1017/S0373463312000082).
- [31] P. D. Groves, Z. Jiang, M. Rudi, P. Strode, A portfolio approach to NLOS and multipath mitigation in dense urban areas, in: Proceedings of the 26th International Technical Meeting of The Satellite Division of the Institute of Navigation, The Institute of Navigation, 2013.
- [32] J. Spilker Jr, Foliage attenuation for land mobile users, *Global Positioning System: Theory and Applications* 1 (1996) 569–583.
- [33] P. Sigrist, P. Coppin, M. Hermy, Impact of forest canopy on quality and accuracy of GPS measurements, *International Journal of Remote Sensing* 20 (18) (1999) 3595–3610. doi:[10.1080/014311699211228](https://doi.org/10.1080/014311699211228).
- [34] J. Tuek, J. Ligo, Forest canopy influence on the precision of location with GPS receivers, *Journal of Forest science* 48 (9) (2002) 399–407. doi:[10.17221/11900-JFS](https://doi.org/10.17221/11900-JFS).
- [35] H.-E. Andersen, T. Clarkin, K. Winterberger, J. Strunk, An accuracy assessment of positions obtained using survey-and recreational-grade global positioning system receivers across a range of forest conditions within the Tanana Valley of interior Alaska, *Western Journal of Applied Forestry* 24 (3) (2009) 128–136. doi:[10.1093/wjaf/24.3.128](https://doi.org/10.1093/wjaf/24.3.128).
- [36] D. N. M. Holden, A. A. Martin, P. M. O. Owende, S. M. Ward, A Method For Relating GPS Performance To Forest Canopy, *International Journal of Forest Engineering* 12 (2) (2001) 51–56. doi:[10.1080/14942119.2001.10702446](https://doi.org/10.1080/14942119.2001.10702446).
- [37] S. D. Danskin, P. Bettinger, T. R. Jordan, C. Cieszewski, A Comparison of GPS Performance in a Southern Hardwood Forest: Exploring Low-Cost Solutions for Forestry Applications, *Southern Journal of Applied Forestry* 33 (1) (2009) 9–16. doi:[10.1093/sjaf/33.1.9](https://doi.org/10.1093/sjaf/33.1.9).
- [38] C. Deckert, P. V. Bolstad, Forest canopy, terrain, and distance effects on global positioning system point accuracy, *Photogrammetric Engineering and Remote Sensing* 62 (3) (1996) 317–321.
- [39] R. Sabatini, T. Moore, C. Hill, S. Ramasamy, Avionics-based GNSS integrity augmentation performance in a jamming environment, in: AIAC16: 16th Australian International Aerospace Congress, Engineers Australia, 2015, pp. 469–479.
- [40] M. Wildemeersch, C. H. Slump, A. Rabbachin, Acquisition of GNSS signals in urban interference environment, *IEEE Transactions on Aerospace and Electronic Systems* 50 (2) (2014) 1078–1091. doi:[10.1109/TAES.2013.120094](https://doi.org/10.1109/TAES.2013.120094).
- [41] J. Spilker, Interference effects and mitigation techniques, *Global Positioning System: Theory and applications* 1 (1996) 717–771.
- [42] R. D. J. V. Nee, Multipath Effects on GPS Code Phase Measurements, *Navigation* 39 (2) (1992) 177–190. doi:[10.1002/j.2161-4296.1992.tb01873.x](https://doi.org/10.1002/j.2161-4296.1992.tb01873.x).
- [43] Z. Jiang, P. D. Groves, NLOS GPS signal detection using a dual-polarisation antenna, *GPS Solutions* 18 (1) (2014) 15–26. doi:[10.1007/s10291-012-0305-5](https://doi.org/10.1007/s10291-012-0305-5).
- [44] K. Wesson, T. Humphreys, Hacking drones, *Scientific American* 309 (5) (2013) 54–59.
- [45] K. Hartmann, K. Giles, UAV exploitation: A new domain for cyber power, in: 2016 8th International Conference on Cyber Conflict (CyCon), 2016, pp. 205–221. doi:[10.1109/CYCON.2016.7529436](https://doi.org/10.1109/CYCON.2016.7529436).
- [46] C. Kanellakis, G. Nikolakopoulos, Survey on Computer Vision for UAVs: Current Developments and Trends, *Journal of Intelligent & Robotic Systems* 87 (1) (2017) 141–168. doi:[10.1007/s10846-017-0483-z](https://doi.org/10.1007/s10846-017-0483-z).
- [47] A. Al-Kaff, D. Martn, F. Garca, A. d. l. Escalera, J. Mara Armengol, Survey of computer vision algorithms and applications for unmanned aerial vehicles, *Expert Systems with Applications* 92 (2018) 447–463. doi:[10.1016/j.eswa.2017.09.033](https://doi.org/10.1016/j.eswa.2017.09.033).
- [48] Y. Lu, Z. Xue, G.-S. Xia, L. Zhang, A survey on vision-based UAV navigation, *Geo-spatial Information Science* 21 (1) (2018) 21–32. doi:[10.1080/10095020.2017.1420509](https://doi.org/10.1080/10095020.2017.1420509).
- [49] G. Balamurugan, J. Valarmathi, V. P. S. Naidu, Survey on UAV navigation in GPS denied environments, in: 2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES), 2016, pp. 198–204. doi:[10.1109/SCOPES.2016.7955787](https://doi.org/10.1109/SCOPES.2016.7955787).
- [50] J.-P. Sanchez-Rodriguez, A. Aceves-Lopez, A survey on stereo vision-based autonomous navigation for multi-rotor MUAVs, *Robotica* 36 (8) (2018) 1225–1243. doi:[10.1017/S0263574718000358](https://doi.org/10.1017/S0263574718000358).
- [51] Y. Xu, L. Pan, C. Du, J. Li, N. Jing, J. Wu, Vision-based UAVs Aerial Image Localization: A Survey, in: Proceedings of the 2Nd ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery, GeoAI'18, ACM, New York, NY, USA, 2018, pp. 9–18. doi:[10.1145/3281548.3281556](https://doi.org/10.1145/3281548.3281556).
- [52] M. Calonder, V. Lepetit, C. Strecha, P. Fua, BRIEF: Binary Robust Independent Elementary Features, in: K. Daniilidis, P. Maragos, N. Paragios (Eds.), *Computer Vision ECCV 2010*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2010, pp. 778–792. doi:[10.1007/978-3-642-15561-1\\_56](https://doi.org/10.1007/978-3-642-15561-1_56).
- [53] C. Forster, M. Pizzoli, D. Scaramuzza, SVO: Fast semi-direct monocular visual odometry, in: 2014 IEEE International Conference on Robotics and Automation (ICRA), 2014, pp. 15–22. doi:[10.1109/ICRA.2014.6906584](https://doi.org/10.1109/ICRA.2014.6906584).
- [54] D. Nister, O. Naroditsky, J. Bergen, Visual odometry, in: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004., Vol. 1, 2004, pp. I–I. doi:[10.1109/CVPR.2004.1315094](https://doi.org/10.1109/CVPR.2004.1315094).
- [55] M. W. M. G. Dissanayake, P. Newman, S. Clark, H. F. Durrant-Whyte, M. Csorba, A solution to the simultaneous localization and map building (SLAM) problem, *IEEE Transactions on Robotics and Automation* 17 (3) (2001) 229–241. doi:[10.1109/70.938381](https://doi.org/10.1109/70.938381).
- [56] R. Mur-Artal, J. M. M. Montiel, J. D. Tards, ORB-SLAM:

- A Versatile and Accurate Monocular SLAM System, *IEEE Transactions on Robotics* 31 (5) (2015) 1147–1163. doi:10.1109/TRO.2015.2463671.
- [57] G. Klein, D. Murray, Parallel Tracking and Mapping for Small AR Workspaces, in: Proceedings of the 2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality, ISMAR '07, IEEE Computer Society, Washington, DC, USA, 2007, pp. 1–10. doi:10.1109/ISMAR.2007.4538852.
- [58] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, P. Furgale, Keyframe-based visualinertial odometry using nonlinear optimization, *The International Journal of Robotics Research* 34 (3) (2015) 314–334. doi:10.1177/0278364914554813.
- [59] M. Li, A. I. Mourikis, High-precision, consistent EKF-based visual-inertial odometry, *The International Journal of Robotics Research* 32 (6) (2013) 690–711. doi:10.1177/0278364913481251.
- [60] M. Bloesch, S. Omari, M. Hutter, R. Siegwart, Robust visual inertial odometry using a direct EKF-based approach, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, pp. 298–304. doi:10.1109/IROS.2015.7353389.
- [61] D. Quick, Global Observer successfully completes first flight, <https://newatlas.com/global-observer-first-flight/16049/> (Aug. 2017).
- [62] S. Leprince, S. Barbot, F. Ayoub, J. Avouac, Automatic and Precise Orthorectification, Coregistration, and Subpixel Correlation of Satellite Images, Application to Ground Deformation Measurements, *IEEE Transactions on Geoscience and Remote Sensing* 45 (6) (2007) 1529–1558. doi:10.1109/TGRS.2006.888937.
- [63] Google earth, <https://www.google.com/earth/> (2020).
- [64] Arcgis online, <https://www.arcgis.com/index.html> (2020).
- [65] A. Couturier, M. A. Akhloufi, Relative visual localization (RVL) for UAV navigation, in: Degraded Environments: Sensing, Processing, and Display 2018, Vol. 10642, International Society for Optics and Photonics, 2018, p. 106420O. doi:10.1117/12.2304901.
- [66] A. Couturier, M. A. Akhloufi, UAV navigation in GPS-denied environment using particle filtered RVL, in: Situation Awareness in Degraded Environments 2019, Vol. 11019, International Society for Optics and Photonics, 2019, p. 110190N. doi:10.1117/12.2518516.
- [67] A. Couturier, M. A. Akhloufi, Conditional probabilistic relative visual localization for unmanned aerial vehicles, in: 2020 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), IEEE, 2020, in press.
- [68] M. Warren, M. Greeff, B. Patel, J. Collier, A. P. Schoellig, T. D. Barfoot, There's No Place Like Home: Visual Teach and Repeat for Emergency Return of Multirotor UAVs During GPS Failure, *IEEE Robotics and Automation Letters* 4 (1) (2019) 161–168. doi:10.1109/LRA.2018.2883408.
- [69] R. Brunelli, Template Matching Techniques in Computer Vision: Theory and Practice, John Wiley & Sons, 2009.
- [70] S. Theodoridis, K. Koutroumbas, Chapter 8 - Template Matching, Academic Press, Boston, 2009, pp. 481 – 519. doi:10.1016/B978-1-59749-272-0.50010-4.
- [71] G. J. Van Denalen, D. P. Magree, E. N. Johnson, Absolute localization using image alignment and particle filtering, in: AIAA Guidance, Navigation, and Control Conference, 2016, p. 0647. doi:10.2514/6.2016-0647.
- [72] A. Yol, B. Delabarre, A. Dame, J.-E. Dartois, E. Marchand, Vision-based absolute localization for unmanned aerial vehicles, in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 3429–3434. doi:10.1109/IROS.2014.6943040.
- [73] J. Lewis, Fast template matching, *Vision Interface* 95 (1994) 15–19.
- [74] S. Thrun, Particle Filters in Robotics, in: Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence, UAI'02, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2002, pp. 511–518.
- [75] Bing maps, <https://www.bing.com/maps> (2020).
- [76] D. P. Magree, E. N. Johnson, A monocular vision-aided inertial navigation system with improved numerical stability, in: AIAA Guidance, Navigation, and Control Conference, 2015, p. 0097. doi:10.2514/6.2015-0097.
- [77] Precision Agriculture - RMAX, <https://www.yamahamotorsports.com/motorsports/pages/precision-agriculture-rmax> (2016).
- [78] T. M. Cover, J. A. Thomas, Elements of Information Theory, John Wiley & Sons, 2012.
- [79] R. M. Gray, Entropy and Information Theory, Springer Science & Business Media, 2011.
- [80] X. Wan, J. Liu, H. Yan, G. L. K. Morgan, Illumination-invariant image matching for autonomous UAV localisation based on optical sensing, *ISPRS Journal of Photogrammetry and Remote Sensing* 119 (2016) 198–213. doi:10.1016/j.isprsjprs.2016.05.016.
- [81] C. Kuglin, The phase correlation image alignment method, in: Proceedings of the International Conference on Cybernetics and Society, 1975, pp. 163–165.
- [82] H. Xiang, L. Tian, Method for automatic georeferencing aerial remote sensing (RS) images from an unmanned aerial vehicle (UAV) platform, *Biosystems Engineering* 108 (2) (2011) 104–113. doi:10.1016/j.biosystemseng.2010.11.003.
- [83] B. Patel, Visual localization for uavs in outdoor gps-denied environments (Nov. 2019).
- [84] G. Pascoe, W. P. Maddern, P. Newman, Robust direct visual localisation using normalised information distance., in: BMVC, 2015, pp. 70–1. doi:10.5244/C.29.70.
- [85] D. R. S. Saputro, P. Widyaningsih, Limited memory broyden-fletcher-goldfarb-shanno (l-bfgs) method for the parameter estimation on geographically weighted ordinal logistic regression model (gwolr), in: AIP Conference Proceedings, Vol. 1868, AIP Publishing LLC, 2017, p. 040009. doi:10.1063/1.4995124.
- [86] C. Harris, M. Stephens, A Combined Corner and Edge Detector, in: Proceedings of the Alvey Vision Conference 1988, Alvey Vision Club, Manchester, 1988, pp. 23.1–23.6.
- [87] E. Rosten, T. Drummond, Machine Learning for High-Speed Corner Detection, in: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision ECCV 2006, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2006, pp. 430–443. doi:10.1007/11744023\_34.
- [88] E. Rosten, R. Porter, T. Drummond, Faster and Better: A Machine Learning Approach to Corner Detection, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (1) (2010) 105–119. doi:10.1109/TPAMI.2008.275.
- [89] D. G. Lowe, Distinctive image features from scale-invariant keypoints, *International Journal of Computer Vision* 60 (2) (2004) 91–110. doi:10.1023/B:VISI.0000029664.99615.94.
- [90] B. S. Seema, K. Hemanth, V. P. S. Naidu, Geo-Registration of Aerial Images using RANSAC Algorithm, in: NCTAESD-2014, Vemana Institute of Technology, Bangalore, 2014, pp. 1–5.
- [91] K. C. Saranya, V. P. S. Naidu, V. Singhal, B. M. Tanuja, Application of vision based techniques for UAV position estimation, in: 2016 International Conference on Research Advances in Integrated Navigation Systems (RAINS), 2016, pp. 1–5. doi:10.1109/RAINS.2016.7764392.
- [92] M. A. Fischler, R. C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, *Communications of the ACM* 24 (6) (1981) 381–395. doi:10.1145/358669.358692.
- [93] H. Bay, T. Tuytelaars, L. Van Gool, SURF: Speeded Up Robust Features, in: A. Leonardis, H. Bischof, A. Pinz (Eds.), Computer Vision ECCV 2006, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2006, pp. 404–417. doi:10.1007/11744023\_32.
- [94] Google maps platform, <https://cloud.google.com/maps-platform/> (2020).
- [95] M. Shan, F. Wang, F. Lin, Z. Gao, Y. Z. Tang, B. M. Chen,

- Google map aided visual navigation for UAVs in GPS-denied environment, in: 2015 IEEE International Conference on Robotics and Biomimetics (ROBIO), 2015, pp. 114–119. doi:10.1109/ROBIO.2015.7418753.
- [96] N. Dalal, B. Triggs, Histograms of Oriented Gradients for Human Detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, IEEE, San Diego, CA, USA, 2005, pp. 886–893. doi:10.1109/CVPR.2005.177.
- [97] B. K. P. Horn, B. G. Schunck, Determining optical flow, *Artificial Intelligence* 17 (1) (1981) 185–203. doi:10.1016/0004-3702(81)90024-2.
- [98] D. S. Bolme, J. R. Beveridge, B. A. Draper, Y. M. Lui, Visual object tracking using adaptive correlation filters, in: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010, pp. 2544–2550. doi:10.1109/CVPR.2010.5539960.
- [99] H. P. Chiu, A. Das, P. Miller, S. Samarasekera, R. Kumar, Precise vision-aided aerial navigation, in: 2014 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2014, pp. 688–695. doi:10.1109/IROS.2014.6942633.
- [100] H. Chiu, S. Williams, F. Dellaert, S. Samarasekera, R. Kumar, Robust vision-aided navigation using Sliding-Window Factor graphs, in: 2013 IEEE International Conference on Robotics and Automation, 2013, pp. 46–53. doi:10.1109/ICRA.2013.6630555.
- [101] M. Kaess, H. Johannsson, R. Roberts, V. Ila, J. J. Leonard, F. Dellaert, iSAM2: Incremental smoothing and mapping using the Bayes tree, *The International Journal of Robotics Research* 31 (2) (2012) 216–235. doi:10.1177/0278364911430419.
- [102] M. Mantelli, D. Pittol, R. Neuland, A. Ribacki, R. Maffei, V. Jorge, E. Prestes, M. Kolberg, A novel measurement model based on abBRIEF for global localization of a UAV over satellite images, *Robotics and Autonomous Systems* 112 (2019) 304–319. doi:10.1016/j.robot.2018.12.006.
- [103] A. Masselli, R. Hanten, A. Zell, Localization of Unmanned Aerial Vehicles Using Terrain Classification from Aerial Images, in: Intelligent Autonomous Systems 13, Advances in Intelligent Systems and Computing, Springer International Publishing, 2016, pp. 831–842. doi:10.1007/978-3-319-08338-4\_60.
- [104] E. Rublee, V. Rabaud, K. Konolige, G. Bradski, Orb: An efficient alternative to sift or surf, in: 2011 International Conference on Computer Vision, 2011, pp. 2564–2571. doi:10.1109/ICCV.2011.6126544.
- [105] L. Breiman, Random Forests, *Machine Learning* 45 (1) (2001) 5–32. doi:10.1023/A:1010933404324.
- [106] M. Shan, A. Charan, Google map referenced UAV navigation via simultaneous feature detection and description, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015.
- [107] F. Tombari, L. Di Stefano, Interest Points via Maximal Self-Dissimilarities, in: D. Cremers, I. Reid, H. Saito, M.-H. Yang (Eds.), *Computer Vision – ACCV 2014*, Lecture Notes in Computer Science, Springer International Publishing, 2015, pp. 586–600. doi:10.1007/978-3-319-16808-1\_39.
- [108] E. Shechtman, M. Irani, Matching Local Self-Similarities across Images and Videos, in: 2007 IEEE Conference on Computer Vision and Pattern Recognition, IEEE, Minneapolis, MN, USA, 2007, pp. 1–8. doi:10.1109/CVPR.2007.383198.
- [109] P. F. Alcantarilla, A. Bartoli, A. J. Davison, KAZE Features, in: A. Fitzgibbon, S. Lazebnik, P. Perona, Y. Sato, C. Schmid (Eds.), *Computer Vision – ECCV 2012*, Lecture Notes in Computer Science, Springer Berlin Heidelberg, 2012, pp. 214–227. doi:10.1007/978-3-642-33783-3\_16.
- [110] P. Alcantarilla, J. Nuevo, A. Bartoli, Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces, in: Proceedings of the British Machine Vision Conference 2013, British Machine Vision Association, Bristol, 2013, pp. 13.1–13.11. doi:10.5244/C.27.13.
- [111] S. Leutenegger, M. Chli, R. Y. Siegwart, BRISK: Binary Robust invariant scalable keypoints, in: 2011 International Conference on Computer Vision, IEEE, Barcelona, Spain, 2011, pp. 2548–2555. doi:10.1109/ICCV.2011.6126542.
- [112] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2012, pp. 1097–1105. doi:10.1145/3065386.
- [113] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, *International Journal of Computer Vision (IJCV)* 115 (3) (2015) 211–252. doi:10.1007/s11263-015-0816-y.
- [114] A. Nassar, K. Amer, R. ElHakim, M. ElHelw, A deep cnn-based framework for enhanced aerial imagery registration with applications to uav geolocation, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 1513–1523. doi:10.1109/CVPRW.2018.00201.
- [115] A. Nassar, M. ElHelw, Aerial imagery registration using deep learning for uav geolocation, in: *Deep Learning in Computer Vision: Principles and Applications*, 1st Edition, CRC Press, 2020, Ch. 7, pp. 183–210.
- [116] H. Goforth, S. Lucey, GPS-Denied UAV Localization using Pre-existing Satellite Imagery, in: 2019 International Conference on Robotics and Automation (ICRA), IEEE, Montreal, QC, Canada, 2019, pp. 2974–2980. doi:10.1109/ICRA.2019.8793558.
- [117] K. Amer, M. Samy, R. ElHakim, M. Shaker, M. ElHelw, Convolutional Neural Network-Based Deep Urban Signatures with Application to Drone Localization, in: 2017 IEEE International Conference on Computer Vision Workshops (ICCVW), 2017, pp. 2138–2145. doi:10.1109/ICCVW.2017.250.
- [118] K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, arXiv:1409.1556 [cs].
- [119] O. Ronneberger, P. Fischer, T. Brox, U-net: Convolutional networks for biomedical image segmentation, in: *International Conference on Medical image computing and computer-assisted intervention*, Springer, 2015, pp. 234–241. doi:10.1007/978-3-319-24574-4\_28.
- [120] M.-K. Hu, Visual pattern recognition by moment invariants, *IRE transactions on information theory* 8 (2) (1962) 179–187.
- [121] A. Marcu, D. Costea, E. Slusanschi, M. Leordeanu, A Multi-Stage Multi-Task Neural Network for Aerial Scene Interpretation and Geolocalization, arXiv:1804.01322 [cs].
- [122] F. Yu, V. Koltun, Multi-Scale Context Aggregation by Dilated Convolutions, arXiv:1511.07122 [cs].
- [123] P. J. Besl, N. D. McKay, Method for registration of 3-D shapes, in: *Sensor Fusion IV: Control Paradigms and Data Structures*, Vol. 1611, International Society for Optics and Photonics, 1992, pp. 586–607. doi:10.1117/12.57955.
- [124] Openstreetmap, <https://www.openstreetmap.org/> (2020).
- [125] C. Wang, H. K. Galoogahi, C. Lin, S. Lucey, Deep-LK for Efficient Adaptive Object Tracking, in: 2018 IEEE International Conference on Robotics and Automation (ICRA), 2018, pp. 627–634. doi:10.1109/ICRA.2018.8460815.
- [126] C.-H. Chang, C.-N. Chou, E. Y. Chang, CLKN: Cascaded Lucas-Kanade Networks for Image Alignment, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2213–2221. doi:10.1109/CVPR.2017.402.
- [127] S. Baker, I. Matthews, Lucas-Kanade 20 Years On: A Unifying Framework, *International Journal of Computer Vision* 56 (3) (2004) 221–255. doi:10.1023/B:VISI.0000011205.11775.fd.
- [128] Earth Explorer, <https://earthexplorer.usgs.gov/> (2020).
- [129] J. Engel, V. Koltun, D. Cremers, Direct Sparse Odometry, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40 (3) (2018) 611–625. doi:10.1109/TPAMI.2017.

2658577.

- [130] H. Alismail, B. Browning, S. Lucey, Photometric Bundle Adjustment for Vision-Based SLAM, in: S.-H. Lai, V. Lepetit, K. Nishino, Y. Sato (Eds.), Computer Vision ACCV 2016, Lecture Notes in Computer Science, Springer International Publishing, 2017, pp. 324–341. doi:10.1007/978-3-319-54190-7\_20.
- [131] senseFly - Datasets, <https://www.sensefly.com/education/datasets> (2020).
- [132] M. Schleiss, Translating aerial images into street-map representations for visual self-localization of uavs, ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences 4213 (2019) 575–580. doi:10.5194/isprs-archives-XLII-2-W13-575-2019.
- [133] P. Isola, J.-Y. Zhu, T. Zhou, A. A. Efros, Image-To-Image Translation With Conditional Adversarial Networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 1125–1134. doi:10.1109/CVPR.2017.632.
- [134] M. Mirza, S. Osindero, Conditional Generative Adversarial Nets, arXiv:1411.1784 [cs, stat].
- [135] G. Conte, P. Doherty, Vision-Based Unmanned Aerial Vehicle Navigation Using Geo-Referenced Information, EURASIP Journal on Advances in Signal Processing 2009 (1). doi:10.1155/2009/387308.
- [136] P. H. S. Torr, A. Zisserman, MLESAC: A New Robust Estimator with Application to Estimating Image Geometry, Computer Vision and Image Understanding 78 (1) (2000) 138–156. doi:10.1006/cviu.1999.0832.
- [137] S. Anderson, T. D. Barfoot, Full STEAM ahead: Exactly sparse gaussian process regression for batch continuous-time trajectory estimation on SE(3), in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2015, pp. 157–164. doi:10.1109/IROS.2015.7353368.
- [138] F. Fidler, J. Wilcox, Reproducibility of scientific results, in: E. N. Zalta (Ed.), The Stanford Encyclopedia of Philosophy, winter 2018 Edition, Metaphysics Research Lab, Stanford University, 2018.
- [139] S. M. Easterbrook, Open code for open science?, Nature Geoscience 7 (2014) 779. doi:10.1038/ngeo2283.
- [140] Embedded Solutions for Drones & UAVs from NVIDIA Jetson, <https://www.nvidia.com/en-us/autonomous-machines/uavs-drones-technology/> (2020).

Moulay Akhloufi received his M.Sc. and a Ph.D. in Electrical Engineering from Ecole Polytechnique of Montreal and Laval University (Canada), respectively. He also holds an MBA from Laval University. He is currently Professor in the University of Moncton (Canada) and Head of the Perception, Robotics, and Intelligent Machines Group. Before Joining the University of Moncton in 2016, he worked in the industry and technology transfer in the areas of computer vision and robotics for over 15 years. He is a Senior Member of IEEE and member of SPIE.

Andy Couturier received a B.Sc.A. degree in computer science in 2015, a B.Sc. in mathematics in 2017, and a M.Sc. in computer science in 2019 from Université de Moncton, Canada. He is currently a Ph.D. candidate in applied science at Université de Moncton, Canada. His interests include computer vision, deep learning, unmanned aerial vehicles, and computationally efficient algorithm for robotic applications.





**Declaration of interests**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

