# Visual–Inertial Navigation Systems for Aerial Robotics: Sensor Fusion and Technology

Fendy Santoso, Matthew A. Garratt, and Sreenatha G. Anavatti

***Abstract*—In this paper, we comprehensively discuss the current progress of visual–inertial (VI) navigation systems and sensor fusion research with a particular focus on small unmanned aerial vehicles, known as microaerial vehicles (MAVs). Such fusion has become very topical due to the complementary characteristics of the two sensing modalities. We discuss the pros and cons of the most widely implemented VI systems against the navigational and maneuvering capabilities of MAVs. Considering the issue of optimum data fusion from multiple heterogeneous sensors, we examine the potential of the most widely used advanced state estimation techniques (both linear and nonlinear as well as Bayesian and non-Bayesian) against various MAV design considerations. Finally, we highlight several research opportunities and potential challenges associated with each technique.**

***Note to Practitioners*—Robotic aircraft have been widely implemented to improve safety, efficiency, and productivity (e.g., agriculture, law enforcement, building inspections, and so on). As a part of its autonomous navigation system, this review aims to address several aspects of VI navigation systems both from data fusion and technological perspectives.**

***Index Terms*—Microaerial vehicles (MAVs), sensor fusion, visual–inertial (VI) navigation systems.**

## I. Introduction

AUTONOMOUS flying robots, colloquially known as unmanned aerial vehicles (UAVs) or drones, have been widely developed for military and civilian applications, such as aerial surveillance, border protection, battle damage assessment, law enforcement, building inspection, cinematography, search and rescue, remote inspection, and volcano observations [1]–[5]. Numerous innovative research groups are committed to this fast growing field across the globe. The smart skies project conducted by the Australian Research Centre for Aerospace Automation [6] and the Aerobatic Airplane project conducted by the Swiss Federal Institute of Technology, ETH Zurich, Zurich [7] are two examples of research projects from academia.

Supported by recent advancements in microelectromechanics systems (MEMS) as well as control and communication technologies, the development of UAVs has led to the birth of aircraft with substantially smaller physical dimensions, i.e., microaerial vehicles (MAVs), as reported in [1]. Our research group in the University of New South Wales Canberra, Australia, has focused on the development of robust visual/inertial navigation algorithm for rotorcraft MAVs, as shown in Fig. 1 [8].

Fig. 1. VI navigation systems to guide our Pelican quadrotor on an internationally recognized "**H**" landing symbol. The system can be implemented to land a UAV on the deck of a moving ship. The UNSW Canberra Pelican quadrotor is equipped with a ToF CAmBoard nanocamera to compute range at each pixel with frame rates up to 100 frames/s [8].

There is a great deal of benefits derived from operating MAVs in both indoor and outdoor environments for civilian and military domains. MAVs can be employed in critically dangerous missions without risking operators while operating in 3D spaces inaccessible to manned aircraft or conventional UAVs, which are larger in size. MAVs can be used to support law enforcement, e.g., surveillance, infrastructure maintenance and management systems [9], precision farming [10], and aerial photography and cinematography [11]. In addition, MAVs have substantially lower costs than regular aircraft or larger UAVs as they possess minimal carbon footprint per kilometer flown, low noise emissions, and minimum operational downtime [4]. Thus, it is essential to equip MAV with a high degree of autonomy including the ability to maneuver among obstacles and to follow a predefined trajectory. However, one of the major challenges with an MAV is related to its limited mechanical and computational payload as well as limited telemetry bandwidth [12].

Much of the work on MAV guidance and control is inspired by biological systems [12]–[15], that is, the ability of insects with tiny brains (less than one-tenth of a milligram) to hover and to fuse multiple sources of information (e.g., vision, smell, and balance), in order to guide its movement. Hence, one of the fundamental research challenges is related to the development of optimum data fusion among multiple heterogeneous sensors to robustly extract meaningful information in order

Manuscript received August 23, 2015; revised November 20, 2015 and April 7, 2016; accepted June 14, 2016. This paper was recommended for publication by Associate Editor B. Vogel-Heuser and Editor M. Wang upon evaluation of the reviewers' comments. This work was supported by The University of New South Wales, Canberra, Australia.

The authors are with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT 2612, Australia (e-mail: f.santoso@adfa.edu.au; m.garratt@adfa.edu.au; s.anavatti@adfa.edu.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TASE.2016.2582752

to provide autonomous MAV navigation systems. The aim is to develop low-cost, reliable, computationally efficient data fusion algorithms, which can be executed using the onboard processing capacity [12]. For instance, Green *et al.* [16] adopt the behaviors of insects to develop an MAV sensor suite.

Although GPS is quite useful for outdoor navigation, there are several issues associated with it. Multipath, urban canyons, jamming, and selective availability can all make GPS systems unreliable. Given the limited functionality of GPS systems within indoor and urban environments, the UAV navigation systems need to be self-contained, robust, and sufficiently accurate to achieve autonomy. Other sensor technologies, such as radar, radio frequency identification (RFID), infrared (IR) sensors, and scanning laser rangefinders (LRFs), may be impractical for MAVs considering their payload limitations [12]. Visual–inertial (VI) navigation systems turn out to become a good candidate to address this situation, since they are self-contained, passive, and can provide a good balance between cost and accuracy.

This paper mainly looks at the technical aspects of the VI navigation of MAVs and its contribution can be elaborated as follows. First, we believe that this review is the first in the literature that address the topic of VI navigation specifically designed for MAVs. This review also provides a comprehensive discussion about several sensor technologies relating to VI navigation systems. Next, this review discusses the pros and cons of various estimation techniques, which are required to optimally fuse multiple information from heterogeneous sensors. We also discuss current state-of-the-art applications to illustrate the feasibility of the algorithms. Finally, we also discuss several important considerations in developing robust VI navigation systems.

This paper is organized as follows. Section II discusses data fusion techniques including discussion relating to state estimators suitable for robotics, while Section III presents the pros and cons of current visual sensor technologies. Section IV depicts the state-of-the-art applications of VI navigation systems for UAVs, followed by comprehensive discussions about potential research avenues in Section V. Finally, Section VI concludes this paper.

## II. Data Fusion: Architectures and Methods

In robotics, we are often faced by the problem of finding the best estimate from noisy and biased measurements, with a specific aim to filter out the noise or to estimate the value of the hidden states that may not be directly available for measurements. In this context, state estimators have been widely implemented to support the development of highly reliable navigation systems.

More specifically, data fusion for heterogeneous sensors, also widely known as multisensor data fusion, is the process of obtaining a robust and reliable description of the dynamic behaviors of the systems egomotion to assist with localization and navigation capabilities of the systems. Interested readers in this topic may refer to [17]–[23], to name a few.

We formulate our high-level problem statement as how could one develop an autonomous navigation solution, which is capable of optimally integrating information obtained from multiple heterogeneous sensors, with minimum complexity in real time. We shall discuss some important issues in data fusion techniques from architectural and methodological aspects as follows.

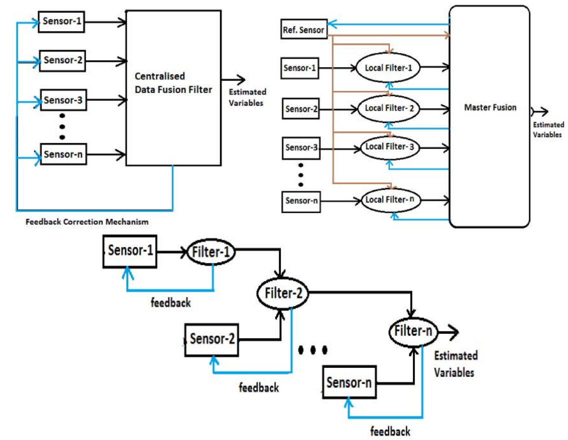


Fig. 2. Most commonly implemented architectures of information fusion for multiple heterogeneous sensors [24]. Top-left: centralized fusion. Top-right: federated fusion. Bottom: cascaded fusion. Both federated and cascaded fusion can also be regarded as forms of distributed approaches.

### A. Data Fusion Architectures

Although there are several data fusion architectures, the most straightforward one is centralized data fusion (see Fig. 2). This happens when all measurements from sensors are directly fed to a master filter. Although this method is reported to be able to achieve optimal estimates of the egomotions of an MAV, the system can be reasonably complex and time-consuming due to its computational demand [24]. A good example of centralized data fusion can be found by Zhan *et al.* [25] where they implemented particle filters for UAV target tracking. Zhai *et al.* [26] proposed a similar system for collaborative sensor networks. Their research indicates that their approach is robust and effective to reduce the computational burden. The authors implemented the Monte Carlo methods to accommodate complex motion patterns in target tracking.

One way to overcome the limitation of centralized data fusion (e.g., speed, reliability, and complexity) is to consider a cascaded data fusion architecture. In this scenario, the outputs of one filter will be fed to the subsequent filter as inputs. This will include the estimates of the system states and their covariance. The main benefit of this architecture is related to improved accuracy as well as the ability of performing in-flight calibration. However, access to information relating to covariance may not be available and the processing time may deteriorate. Since the cascaded architectures employ a series configuration, the performance of the primary filter plays a more significant role in determining the accuracy of the whole system as the error will accumulate throughout the processes. This method also relies on the calculation of the covariance of the primary filter to suppress the propagation of the error [24].

Owing to the benefit of the parallel architecture, Carlson [27] proposed a federated filter architecture where all parallel processing are considered local [see Fig. 2 (top-right)]. The system is connected with a common reference system to obtain local estimates while global estimates can be obtain by fusing these local estimates in conjunction to their weighted covariance in a master filter. Interested readers on the development of cutting-edge distributed fusions for target tracking may refer to [17], [24], [28], and [29].

Unlike centralized architecture, distributed data fusion has no fixed or standard model—cascaded and federated filters can also be regarded as a subset of distributed fusion models. It comprises of two main stages, that is, measurement and state fusions. While in state fusion, local estimations are fused in a master filter to obtain global estimates; in measurement fusion, a set of sensor measurements are fused using filter banks to acquire global estimates. This method turns out to be the most flexible scheme in designing an MAV navigation system [24].

Some examples of decentralized data fusion in UAV navigation systems can be found by Kim *et al.* [30] and Sukkarieh *et al.* [31], where they proposed autonomous navigation systems without the need for *a priori* infrastructures (e.g., GPS, beacons, or preloaded maps of the area). The systems have a built-in capacity to map the building in a distributed manner. The authors employ extended Kalman filters (EKFs) as state estimators to fuse information from multiple sensors. The architecture of the data fusion technique is characterized by the absence of central data fusion, common communication facility, as it is performed on the node-to-node basis.

### B. Data Fusion Methods Based on Filtering

In this section, we shall briefly discuss some well-known state estimators applied in robotics. To highlight the principles of three well-known state estimation concepts (i.e., filtering, prediction, and smoothing), one can employ the following mathematical formulation. Given two signals $x$ and $y$, the task of the filter $F$ is to estimate the signal $y$, such that $\hat{y} = Fx$ holds. In other words, given $y(t) = s(t + \tau)$ and $x(t) = s(t) + d(t)$, where $s$ denotes the signal of interest and $d$ indicates disturbance, the task is called predicting, if $\tau > 0$; it is considered filtering, if $\tau = 0$; and smoothing, if $\tau < 0$ [32].

*1) Kalman Filter (KF) and Extended Kalman Filter:* KF, known as the minimum-variance state estimator, has been widely used for more than 50 years to compute optimal estimations to a somewhat restricted classes of linear Gaussian problems [17]. Originally developed to estimate the behaviors of linear systems under Gaussian noise, the KF has been widely known as an optimal filter that minimizes the mean square error (MSE) of the estimated parameters. Using the KF, the measurement data are not only cleaned up from noise, but also they can be optimally projected onto the state estimates.

The KF has been extensively used as a powerful state estimator in various scientific fields, such as robotics, navigation and indoor positioning systems, flight control systems, weather forecasting, and many more. The attraction of the KF is because of its simplicity in designing, coding, and fine-tuning, making it convenient for real-time processing in many engineering applications [29]. However, the KF suffers from several technical issues that may limit its practical usefulness, namely, poor robustness in the face of nonlinear mathematical models and uncertainties introduced by an ill-conditioned covariance matrix as well as inaccuracies in modeling. To overcome this issue, several modifications of the KFs have been proposed in the literature.

According to Chen *et al.* [51], today, there have been around 20 types of KFs (e.g., ensemble KF [19], EKF [22], extended information filter [23], multistate constraint KF (MSCKF) [39], and unscented KF (UKF) [52]). In this section, we will limit our discussion on some of the most commonly used KFs in visual/inertial navigation systems (INSs).

Owing to the principle of the KF, the idea of the EKF is to estimate the nonlinear system with its linearized approximations about the current mean and covariance by employing the first-order Taylor series. As a nonlinear extension of the KF, the EKF has received considerable attention for its practical usefulness as it has become the most widely used version of the KF applied in robotics in the past few decades [17], [43]. Based on the linearization concept, the EKF employs Jacobian matrices to derive the state-transition and observation matrices [22], [53].

*2) Simultaneous Localization and Mapping:* The simultaneous localization and mapping (SLAM) system aims to construct and update the map of an unknown environment while keeping track of the position and orientation of the observer. Researchers have widely implemented the SLAM techniques to achieve accurate performance in the face of highly nonlinear optimization tasks. Although the main advantage of SLAM is its accuracy, the major limitation of this technique is its computational complexity. Performing vision-based SLAM in an environment with thousands of landmarks is a challenging task.

To date, there are two popular solutions of SLAM, namely, the EKF and the particle filter. While EKF-SLAM can be very useful to deal with nonlinear systems, this technique may not lead to optimal results, since the system propagates linearization errors in the estimation process. Another problem with the applications of the EKF-SLAM is its computational intensive nature [22], [44], [52]. There are basically two fundamental approaches in SLAM, namely, batch and online filtering. While batch solution works by giving all data and solutions upfront (offline), online filtering provides an autonomous solution by computing the states of the system at each sampling time as more data are being received.

*3) Unscented Kalman Filter:* The propagation of a Gaussian random variable plays an important role in determining the accuracy of the KF. As the state variables are represented with Gaussian distribution, the first-order linearization of the nonlinear systems can introduce a reasonable amount of errors in true posterior of the mean and covariance values, which can lead to suboptimal performance or even the growth of the estimation error as the process diverges.

Motivated by the idea to overcome the limitation of the EKF, the UKF works by employing an unscented transform to determine the minimum set of sample points around the mean to avoid poor estimation performance, especially when the system models are highly nonlinear. The term unscented has no special technical meaning, and its origin is somehow related to the word absurd [54].

The UKF employs a deterministic sampling approach [52]. While the state distribution is approximated using a Gaussian distribution, the minimum set of carefully chosen points is now employed to avoid the divergence of the true mean and covariance of the Gaussian probability density function from its original points. As a result, these points can be propagated through the true nonlinear system in order to achieve a more accurate system—the posterior mean and covariance are accurately estimated up to the third order of Taylor series expansion, as opposed to the EKF, which only achieves the first-order estimation. Nevertheless, the UKF requires the same computational complexity as the EKF.

*4) $H_\infty$ Filtering Techniques:* The concept of the $H_\infty$ filter was originally used in robust control to minimize the worst case scenario of the estimation error by accommodating large amounts of uncertainties in system and noise models rather than the covariance of the estimation errors as in KFs. While the $H_\infty$ filter is optimal in the sense of minimizing the infinity norm of the gain between a set of disturbance inputs and the estimation error, the KF aims to minimize the mean square gain between the disturbance and the error of estimator. Given uncertainties in modeling, the $H_\infty$ filter clearly shows some benefits over the KF, which heavily relies on the mathematical model of the system and the knowledge of the noise model. Abdelkrim *et al.* [48] proposed the use of the robust $H_\infty$ filter to address the limitations of the EKF related to the initialization problem as the linearization errors could severely degrade the performance of the MAV localization estimates. The authors demonstrated that without any assumption on the process or/and noise, especially for strong nonlinearity, $H_\infty$ nonlinear filter performed reasonably better compared with the EKF counterparts.

*5) Particle Filters (Sequential Monte Carlo Methods):* As there has been a surging interest in nonlinear and non-Gaussian filtering, the sequential Monte Carlo estimations or collectively referred to as particle filters have become the most widely used method for stochastic estimation problems [17], [43], [55]. As a nonlinear filtering based on the Bayesian techniques (prior results shall determine the probability of the current and future events as the prediction is continuously revised in light of new evidence), we can consider particle and KFs as virtually two opposite ends of the spectrum of Bayesian estimators with some potential benefits outlined as follows.

First, the particle filters (PF) offer an array of benefits, compared with EKF, especially for systems with highly nonlinear mathematical models operating under non-Gaussian noise measurements. This will lead to a more realistic approach, since most systems, in practice, are inherently nonlinear; and most errors in measurements are inherently non-Gaussian. PF can work reasonably better under those circumstances compared with the conventional KF [17], [29], [56], [57]. In fact, the more nonlinear and the more non-Gaussian the noise, the more potential the benefits offered by PF as opposed to the KF [29].

Second, unlike the EKF/UKFs, PF can ease the mathematical complexity in the absence of complex linearization of the mathematical models. Instead, one now can employ a cloud of particles to statistically represent the estimates of the true states. Although, in the past, the real-time applications of the PF were hampered by their demanding computational burdens, this issue is getting progressively less relevant due to the increased speed of computers.

Yet another advantage of employing PF for indoor MAV, compared with the EKF counterparts, is the provision of using *a priori* knowledge regarding the environment known as digital map information to substantially improve accuracy. The designers can assign the weight of zero to particles that cross over a wall of boundary leading to a significant improvement in accuracy [56]–[59]. Although PF has been proven to successfully outperform the EKF, with the proper precautions, it should be noted that the EKF could also outperform PF in a number of situations, e.g., a well-tuned KF that operates under Gaussian noise environments [22].

*6) Rao-Blackwellized Particle Filters and Fast SLAM:* The idea of Rao-Blackwellized particle filter (RBPF) is to achieve the best of both KF and particle filter [60]. When a linear submodel is contained in the state model of nonlinear system, the linear state components can be separated and individually estimated by KFs while the nonlinear state components can be dealt with particle filters. This technique can be expected to achieve superior performance not only when it comes to accuracy under severe fading environments (nonlinear and non-Gaussian noise distribution) but also to achieve a very competitive computational load and convergence rate [61].

The main idea of the Rao-Blackwell method is to offer dimensional reduction that makes use of structure of the models to split the conditional posterior into two different parts when the components of the states can be subdivided into two classes (one that satisfies Gaussian distribution and linear models and the other class is for non-Gaussian distribution). Subsequently, PF can be used to generate the non-Gaussian posterior and KFs are implemented for Gaussian counterparts.

Therefore, although a myriad of filters has been proposed to tackle the problem of distributed fusion of inertial measurement units (IMUs), the RBPFs could become a promising candidate for an optimal fusion, since it will not only provide a new approach that significantly reduces the computational load, but also provides an attractive convergence rate while maintaining the performance of the standard PF when subjected to nonlinearity and non-Gaussian environments [61].

PF also has been used as an alternative for SLAM in addition to EKF. However, the solution from PF may deteriorate, as the dimension of the state steadily increases. The presence of landmarks can substantially enlarge the dimension of the system states, leading to a gradual decrease in the accuracy of PF. Looking back at the classical solutions, one may consider the use of the KFs (e.g., EKF, UKF, etc.) to perform continuous tracking of all variables in one joint distributions. Nevertheless, this will result in a huge covariance matrix and could also reduce the accuracy.

Thus, one possible solution to this problem is to employ a smaller covariance matrix by developing a multimodal or hybrid estimation system, while being scalable $\mathcal{O}(n^2)$, where $n$ is the number of landmarks. This solution leads to the so-called fast-SLAM algorithm, which is a smart tracking solution, owing to the benefits of hybrid KF and particle filter. This method is widely known as an RBPF. The Fast SLAM is a reflection of an RBPF, that is, by employing an individual KF for each PF will eliminate the need to sample the landmark from the whole joint distributions, resulting in faster, accurate, and more robust solutions.

*7) Multistate Constrained Kalman Filter:* MSCKF has been widely implemented in VI navigation systems to perform tightly coupled VI systems over a sliding window of $m$ poses. To determine probabilistic constraints between poses, the system employs all observations available within its sliding window.

Huang *et al.* [38] introduced the concept of consistent estimation for VI navigation systems based on the concept of MSCKF. The authors imposed constraints in both state-transition and observability analysis in determining the Jacobian matrices of the EKF, so that the propagation of the Jacobian remains an appropriate state-transition matrix. Accordingly, the resulting linearized system still inherits the properties of the underlying nonlinear counterparts to satisfy the rule of semigroup property. This can be achieved by constraining the orientation of the error state in global coordinate frame of reference before one can evaluate the propagated Jacobian matrix, instead of the updated state estimates. This way, one can always ensure to achieve correct observability by employing the most accurate (canonical) Jacobian matrix onto the observable subspace in order to avoid the propagation of the false information to retain the consistency of the estimation results. The authors indicate that through a series of Monte Carlo simulations, the proposed algorithm can clearly outperform the regular VI navigation systems when it comes to consistency and accuracy. We summarise the pros and cons of the most widely implemented filtering techniques in robotics in Table I.

### *C. Data Fusion Methods Based on Smoothing*

We also notice that there have been a myriad of powerful estimation methods that have been developed, e.g., iterative nonlinear optimization. Due to their computationally efficient nature, some of these techniques can outperform Bayesian frameworks in a number of situations under proper precautions [20], [62]. To date, there are multiple iterative-based optimization techniques that have been widely implemented in robotics as alternatives to Bayesian filtering approaches.

*1) Sliding Window:* Chiu *et al.* [63] introduced a robust and low latency solution for vision-aided navigation by means of sliding-window factor graphs. In Bayesian estimation, factor graphs are widely known as a graph-based smoothing method. The proposed method extends incremental smoothing methods to deal with the existing states inside a sliding window. The estimation process is performed in two main streams. The first step is rapid smoothing that achieves short-term optimization by estimating the states over a fixed length sliding window given fixed computational cost. Second, a slower and yet fully global smoother is to determine the optimal solution of the nonlinear problem due to loop-closure constraints, which can be regarded as a problem of identifying a previously visited location and a shared map for 3D view. The algorithm leverages on the advantages of smoothers to optimize the 3D landmark map, while minimizing the computational burden. To achieve the optimal solutions, the authors formulated a nonlinear least-square problem to accommodate all received measurements. Unlike traditional linear smoothers, the system can support relinearization and efficient updates.

*2) Fixed-Lag Smoothing:* Smoothers have been widely considered as ones that have the best inherent performance characteristics [64]. One technical advantage of smoothers relative to filters is due to their inherently less complex nature while being able to produce true online estimation. Conceptually, there are three types of smoothers, i.e., fixed-point smoothers, fixed-lag smoothers, as well as fixed-interval smoothers [65]. One of the most useful smoothers is a fixed-lag smoother, since the outcome of the estimation is online, despite there being a small fixed delay [64]. In what follows, we will discuss some state-of-the-art applications of fixed-lag smoothing in robotics.

Dong-Si and Mourikis [66] developed a fixed-lag smoothing algorithm to the maximum *a posteriori* (MAP) estimator to address the issue of accurate motion tracking in autonomous vehicle navigation. The authors employed a set of exteroceptive sensors (e.g., camera and laser scanner) as well as proprioceptive measurements (e.g., odometry and IMU) to estimate the trajectory of the vehicle. The authors focused on processing local information in the absence of loop closing (i.e., localization by means of *a priori* information of the landmarks). An iterative minimization technique and smoothing were performed within the framework of information filtering to deal with the effects of highly nonlinear measurements. Furthermore, the algorithm works by marginalizing out old states to achieve bounded complexity. The major contribution of their research is related to comprehensive analysis of the effects of marginalization on the consistency of the estimator. The EKF-based fixed-lag smoothing is prone to a gradual increase in the linearization error, a crucial issue in the area of motion tracking, where the error can easily accumulate in the absence of *a priori* information (e.g., landmark).

As in [66], one may consider the use of the sliding-window technique to achieve good estimates of currently active state variables. However, discarding old states, as the camera moves in space and new states are being added, may not lead into an optimal solution. This is mainly due to a fixed window assumption, which are employed to bootstrap the estimates, since this method ignore the uncertainty of the discarded states, known as the marginalization issue. Accordingly, the authors proposed a linearization scheme, resulting in substantial improved of accuracy compared with the regular approach. In more details, the authors elaborated the derivation of the marginalization equation, before analyzing the effects of the marginalization on the consistency of the state estimates. In particular, they indicated that the marginalization process

TABLE I
PROS AND CONS OF MOST WIDELY IMPLEMENTED FILTERING-BASED FUSION TECHNIQUES

| No | Methods | Pros | Cons |
|---|---|---|---|
| A1 | Kalman Filter (KF) [18], [23], [30], [34] | Provides an optimum solution for Gaussian noise and linear systems. | Poor robustness in the face of non-linear mathematical models and uncertainties introduced by an ill-conditioned covariance matrix as well as inaccuracy in modelling. |
| A2 | Extended Kalman Filter (EKF) [30], [35] (*e.g* EKF-SLAM) | Intended to estimate the non-linear system model with linearised approximations about the current mean and covariance, Being extensively used (practical usefulness). | The EKF is a heuristic approach and hence there does not exist any theoretical property of convergence and as such there is not any guarantee that the filter works. It also does not lead to optimal results as the system may propagate a reasonable amount of linearisation errors in the estimation process. The EKF-SLAM is expensive in terms of computational complexity as it is quadratic in the number of features in the state vector. |
| A3 | Unscented Kalman Filter (UKF) [36], [19] | Employing unscented transform to determine the minimum set of sample points around the mean to avoid poor estimation performance, especially when the system models are highly non-linear. | Similar computational complexity of the EKF. |
| A4 | Robust Adaptive Kalman Filter (RAKF) [37] | Capable of accommodating large uncertainties such as for fault detection to secure the robustness of the systems. It can still give accurate estimation in the face of the failure of the sensors. | Limited to linear system. |
| A5 | Adaptive Robust Extended Kalman filter (AREKF) [38] | Robustness property, ability to accommodate non-linearity and parameter changes. Flexibility due to its capability of switching between Robust Extended KF mode and Adaptive EKF mode. | More computationally intensive. |
| A6 | Multi-State Constrained Kalman Filter (MSCKF) [39], [40] | Maintaining high level consistency and accuracy by ensuring appropriate state-transition and observability properties to avoid the propagation of false information, linear complexity. | Current research has not addressed issues relating to how to efficiently integrate loop closure for long-term navigation while achieving bounded errors. |
| B1 | Luenberger Observer [41] | Unconstrained, less computational burden. | Sub-optimum estimator, limited to linear systems. |
| B2 | Extended Luenberger Observer [42] | Suitable for non-linear systems, more realistic and precise estimation technique. Straight forward applications to MIMO system. | Linearisation burden, differentiability, that is, based on an assumption of *n*-fold differentiability of the non-linear systems. |
| B3 | Adaptive Luenberger Observer [43] | More realistic approach to accommodate parameter changes in the system, i.e. variations in voltage-to-thrust relationship, adaptive tuning of the parameters of the Luenberger observer. | Limited functionality because of cross-coupling nonliterary from the dynamics of the system, applicable under near hover flight conditions only. |
| C1 | Particle Filter [18], [44], [45] (*e.g* PF-SLAM) | Suitable for linear/non-linear systems, with Gaussian/non Gaussian noise model, Can ease the mathematical complexity due to the absence of the requirement to perform linearisation in the mathematical models, Widely implemented for SLAM. | Computationally intensive, unsuitable for system with limited payload. |
| C2 | Rao-Blackwelised Particle Filter [46], [30] (*e.g* Fast SLAM) | Combining the benefits of Kalman and Particle Filters, offer dimensional reduction based on the structure of the models, widely implemented for fast SLAM. | The derivation of the Jacobian matrices and the linear approximations of nonlinear functions. |
| C3 | Self-Adaptive Particle Filter [47] | Remove the need of overestimating the number of particles by allocating particles in areas of high likelihood, has a new adaptation mechanism to adapt the importance function, similar computational burden to regular PF. | New adaptation algorithm can increase the complexity and processing time of the system. |
| C4 | Evolutionary Particle Filter [48] | It combines a multi-feature fusion method and a genetic evolution mechanism to guarantee robustness and accuracy in the face of challenging environments, i.e. variation in illumination as well as interference of similar targets. | It is more computationally intensive to accommodate both cloud of particles and optimisation technique by means of the genetic algorithm. |
| D1 | Robust $H_\infty$ Filter [49], [30] | No requirement to have *a-priori* knowledge about noise model of the environments, robustness to uncertainties (e.g. process and measurement noises). | Mathematical and computational complexity. |
| D2 | Extended Robust H-infinity Filter [50] | It can accommodate nonlinear uncertain system with an integral quadratic resulting in robustness against disturbance and parameter variations. | Approximated solution, linearisation burden. |
| D3 | Robust $H_\infty$ - Particle Filter [51] | Combine the advantages of particle and $H_\infty$ Filters, can be tuned to compromise between accuracy and robustness by adjusting disturbance attenuation factor. | Computationally intensive. |

can cause the actual error covariance to be larger than one reported by the estimator due to the presence of artificial information. Accordingly, the authors introduced a simple modification in the choice of the linearization points to avoid the propagation of such information in the estimator. It is important to consider the observability properties of the system when choosing linearization points to ensure the linearized model inherits the same properties. The authors also noted that their work was the first to address the effects of the marginalization process on the consistency of the estimates. The authors highlighted the efficacy of their algorithm through computer simulation and real-time experiment.

Mottaghi *et al.* [67] addressed the issue of place recognition by means of the measurements taken from noisy odometry and stereo matching under unreliable information of global position. Some places are manually marked and the least-squares optimization techniques are implemented to improve their localization accuracy. Odometry and GPS measurements are used in the beginning and in the end of the trajectory of the robot to learn the location of some arbitrary key places, that is, to search correspondence between the current observation and the special place in database. The outcome is given in the form of a new 6D global measurement with a given mean and covariance. In the absence of a GPS signal, this new measurement is considered to enhance the accuracy of pose estimation. To improve its pose estimate, the fixed lag smoother combines the odometry measurements with the relative location to the key place. Their research indicates that although the system can achieve reasonable accuracy, the current methods indicate some drawbacks. Visiting a key place from a completely different angle may not make the place recognizable. The recognition task relies on the structure and the size of the scene, making it difficult for dynamic environments. The error may accumulate, if the robot is unable to see a key place on its trajectory, and hence, there is no improvement in the pose estimation.

*3) Keyframing:* Nerurkar *et al.* [65] presented a new concept of MAP filtering technique known as constrained keyframe-based localization and mapping (C-KLAM), intended to address the computational intensive nature of a large-scale SLAM while maintaining high-level estimation accuracy. For instance, the EKF, as a minimum MSE estimator, has a computational complexity given by $\mathcal{O}(N^2)$, where $N$ is the number of landmarks in map, at every sampling time, while the worst computational complexity of the batch MAP SLAM is given by $\mathcal{O}(K+N)^3$, where $K$ is the number of robot poses in the system. These computational costs could eventually limit the effectiveness of the systems. The algorithm employs information obtained from both IMU and camera measurements from nonkeyframes to consistently generate pose constraints between the keyframes, which can be achieved by marginalizing the non-keyframes along with the obtained landmarks. Compared with the sliding-window approach, C-KLAM can acquire information from marginalized frames and landmarks to generate accurate solutions, without breaking the sparsity of the information matrix. This technique can offer a less complex solution, since the cost of marginalization, given by $\mathcal{O}(M_r^3)$, is only intended in the number of nonkeyframes, $M_r$ between consecutive keyframes, and in linear relationship to the number of landmarks. This leads to the suitability of the C-KLAM to ensure accurate and consistent long navigation as well as efficient loop closures. This indicates the ability of the system to recognize the previously visited locations, while updating their knowledge. Yet another advantage of this approach is its processing speed, as it only took 4% of the required time for the entire bundle adjustment.

### *D. Loosely and Tightly Coupled Visual-Inertial Systems*

Many researchers have developed loosely coupled VI systems. For instance, Konolige *et al.* [68] incorporated IMU measurement as an external inclinometer as well as a yaw estimator for the stereo vision optimization, while Weiss *et al.* [7] employed vision-only estimates supported by indirect reading from IMU. While loosely coupled VI systems can limit the complexity of the algorithms by ignoring the correlations amongst internal states of heterogeneous sensors, in tightly coupled approaches, all sensor states are carefully considered for the best estimation performance. In addition, loosely coupled VI systems can separately process the measurements from IMU and camera in a front end, before the systems fuse them in a back end [38]. The main drawback of this system is due to decoupling effects, resulting in information loss. Thus, it should be pointed out that the benefits of the tight fusion algorithms are their increased accuracy and robustness, compared with both vision-only and the loosely coupled VI counterparts.

### *E. Observability Analysis*

Observability is an important issue in estimation theory, since it gives an indication about the availability of information for state estimation. A system is observable if one can determine its state at a certain time instant given a finite sequence of its outputs. Considering the potential of the VI navigation system, Jones and Soatto [69] investigated the observability issues of VI system by studying the indistinguishable trajectory of the system under different sensor configurations. Martinelli [70] introduced the concept of continuous symmetry to address the state estimation problem, where the information provided by the sensor is insufficient to carry out the state estimation. The author also introduced a robust and energy efficient solution to calibrate a vision sensor and estimates its systematic error.

Martinelli [71], [72] highlighted the observability properties in VI structure-from-motion (VI-SfM), especially for the case of reduced number of inertial sensors. More specifically, the author addressed the minimum case of single accelerometer without gyroscope in the presence of a single point feature as well as biased measurements and unknown extrinsic camera calibration. Martinelli [72] introduced a simple closed-form solution for VI-SfM. The author showed that depending on the trajectory, number of point features, layout, and number of monocular images, the problem can have either a unique solution, or two distinct solutions, or infinite solutions. This outcome is useful for applications that require solution of the SfM problem using low-cost sensors in the absence of any infrastructure, such as in GPS-denied environments.

Likewise, Weiss *et al.* [7] performed an observability analysis of this sensor fusion based on the Lie's derivative. The authors find that the observability matrix suffers from rank deficiency (rank 20 instead of 24). It was revealed that the position and yaw states of the IMU with respect to global reference point are not observable in the absence of information about the absolute position and yaw orientation. However, the absolute pitch and roll angles of the IMU with respect to global coordinate are observable from the point-of-view IMU's gravity measurement. Based on their analysis, the authors claim that both the visual scale factor and the states of the six degree-of-freedom (DOF) of the intersensor calibration between IMU and camera are also made observable. This clearly indicates the sufficiency for local stabilization of MAV, that is, it is still possible to self-calibrate sensor and perform speed control although it will slowly drift in position and yaw direction.

### *F. Fundamental Measurement Functions*

In this section, we will highlight some well-known measurement techniques implemented in SLAM, namely, large-scale direct (LSD-SLAM), dense tracking and mapping (DTAM), as well as parallel tracking and mapping (PTAM).

*1) LSD-SLAM:* One essential measurement technique in SLAM is LSD-SLAM, which is considered as a new, real-time, and direct monocular SLAM technique. The system employs the intensity of the image for both tracking and mapping. Engel *et al.* [73] advocate a direct monocular SLAM to build a consistent 3D map of the environment that is reconstructed in real-time fashion, and is represented as point clouds. Unlike current approaches, which are purely based on the odometry measurement, the proposed system maintains and tracks the global map of the environment containing pose graph of keyframes supported by its probabilistic semidense map. The main novelty is given in the form of two keyframes that can be aligned by a direct method, while noise on the estimated depth of tracking can be estimated by means of a new probabilistic approach.

Direct image alignment can also be used to locate the position of the camera, while a semidense map can be employed to determine the geometry. Unlike keypoint-based techniques, which can only employ limited information (e.g., small patches), LSD-SLAM can lead to better accuracy and superior robustness, particularly in sparsely textured indoor environments, since the system can employ all information given in an image, including edges [73]. An estimated semidense depth map can be acquired by means of the pose graph of keyframes. Considering a new direct image formulation, the constraints between keyframes, containing an estimated semidense map, can be directly tracked (rigid body motion and scale). This way, substantial scale drift after loop closure and large-scale variation within the same map can be corrected.

*2) DTAM:* Another important feature in SLAM is DTAM by Newcombe *et al.* [74], who developed real-time camera tracking and reconstruction based on the dense of every pixel, instead of feature extractions. The authors employed a handheld red-green-blue (RGB) camera to estimate detailed texture maps to generate a surface patchwork equipped with millions of vertices. Under a nonconvex optimization method, the authors employed a sequence of images obtained from a video stream to improve the quality of photometric data and to minimize a global energy function that is spatially regularized. Energy is defined as the sum of a photogrammetric error data and spatial regularization term. The proposed algorithm utilizes GPU hardware to achieve a real-time performance. This way, the authors clearly demonstrate the superiority of the dense model used in DTAM in terms of tracking performance with rapid motion, compared with one using feature extraction. Considering a dense model of the scene, the authors employed the alignment of the dense image to track the motion of the camera at frame rate. The proposed algorithm is quite useful for real-time scene interaction (e.g., enhanced augmented reality applications). However, since the algorithm was developed under an assumption of constant brightness in all stages of reconstruction and tracking, the system is not robust to variation in local illumination. An alternative to improve the robustness of the system in the face of local and global lighting changes is by integrating normalized cross correlation measure into the objective function [75].

*3) PTAM:* Yet another widely implemented camera tracking system in SLAM is PTAM. An example of a cutting-edge PTAM can be found by Sa *et al.* [76], where they proposed an autonomous navigation system for a low-cost quadcopter in GPS-denied environment based on the concept of a monocular SLAM. The authors addressed the limitation of the monocular camera (e.g., to estimate the scale of the scene) by employing a cost function comprising of a drift-free altitude measurement as well as up-to-scale position estimate, acquired from the camera. The authors evaluated the performance of the system (e.g., scale estimator, state estimator, and the autopilot) by comparing with the ground truth data obtained from the VICON motion capture sensor. The authors employed PTAM for camera pose tracking and employed a metric scale estimation to transfer up-to-scale position into the form of metric. They employed a KF to track four states, including lateral position and velocity. Their research indicates that the proposed system performed reasonably accurate, although its main drawback is related to the fragility of the PTAM system, as it is dependent on the field-of-view (FoV) of the camera. A narrow FoV camera can severely impair the performance of the system, e.g., when the yaw suddenly changes more than $\pi/2$ rad/s. Subsequently, the forward looking camera suffers from ambiguity, especially for the case when insufficient translation leads to inaccuracy in 3D maps and position estimates. It is also worth noting that PTAM has successfully been implemented in the Swarm of Micro Flying Robot project, whose goal is to create a swarm of vision-controlled MAV, capable of performing autonomous flight in a GPS-denied environment [77].

## III. Visual Sensors

In this section, we shall briefly discuss several camera technologies that have been widely implemented as a means of visual navigation system in various robotics platforms, particularly for MAVs.

### A. Monocular and Stereo Cameras

There are several types of cameras that have been widely implemented to support the development of visual SLAM in robotics. This will clearly benefit the development of autonomous and robotic systems, in the absence of external positioning or navigation systems, such as GPS or global navigation satellite system that possess technical limitations in terms of reliability and security.

While stereo cameras can provide reasonably more accurate in-depth information about the structure of the environment, small robotic platforms, like MAVs, may not have sufficient space and mass to accommodate an onboard stereo camera due to their limited dimension and payloads. Thus, monocular vision may become a reasonable solution for MAVs. However, as stated in [78], applying monocular cameras may introduce a new challenge as it can only provide bearing angles of the visual features leading to a new research challenge.

An application of a monocular camera can be found by Chen *et al.* [51], where they proposed the development of a new servo tracking system by means of a monocular camera mounted on a certain UAV to track the relative position of the leading UAV with fixed orientation, i.e., formation flying. The system works by comparing the coordinate points of the leading UAV obtained from the live image and employing projective geometric methods to construct a Euclidian homography of the target tracking. The proposed control system can effectively deal with moving objects or dynamic object in both camera and the target tracking.

Gohl *et al.* [79] employed a stereo camera for autonomous inspection of underground mines. The objective of their

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



research was to fly the UAV to autonomously map vertical shafts in a deep mine. Considering the nature of the environment, the main challenge is to navigate the UAV in a narrow area, around 1.5 m wide. The authors employed a FireFly quadcopter from Ascending Technologies, equipped with an field-programmable gate array-based VI sensor from Skybotix, in addition to four industrial grade Wide Video Graphics Array-CMOS cameras (752 × 480 pixels). An LRF from Hokuyo was mounted on top of the MAV to measure the distance to objects within the range of 5.6 m, with an accuracy of 1 cm. To evaluate the effectiveness of the proposed system in a mine setting, it was tested in a tunnel situated 3.1 km below ground. The authors employed a 3D occupancy grid mapping approach, known as the Octomap to store the sparse points in space.

### B. Optical Flow Sensing

Optical flow vision, known as the apparent motions of brightness across a series of an image [15], relies on the approximation of motion fields computed from a set of image sequences. For instance, for altitude estimation, it is possible to determine the altitude by observing the optical flow, that is, faster flow indicates a low flight altitude and vice-versa, while obstacles can be detected by measuring expansion or divergence of the forward visual field. Although this method of vision turns out to be somewhat noisy, it still becomes a good candidate for MAV navigation, by combining with other sensor, such as inertial. For a downward looking camera, the relation between altitude and velocity for terrain following navigation is approximately given by: $H = V/Q$, where $h$ is the height or the altitude of the MAV, $v$ denotes the horizontal forward velocity, and $Q$ indicates the magnitude of optic-flow vector in response to translational movement [15]. One well-known low-cost quadcopter platform that uses optical flow for its navigation system is the Parrot AR.Drone.

There are two ways of utilizing information from an optical flow sensor. First, optical flow can be used to estimate the range given the reading from a nonvisual sensor (e.g., GPS), with applications in terrain following and centering in a natural canyon [12]. Second, it can be used to estimate the velocity given a range measurement obtained from another sensor, e.g., LRF [12]. However, one cannot expect to obtain both scale and velocity simultaneously using only a single optic-flow sensor, that is to say, if we need speed as the outcome, the height must be known and vice-versa. The major challenge of optic-flow vision is due to the need to scale the system from another sensor to be useful to determine the egomotion [14], [15]. Considering the use of optical flow in an MAV navigation system, it is also necessary to eliminate the effects of rotational movements of the aircraft and only consider the components generated by the translational motion in order to extract range information, resulting in noisy estimation of the range.

To date, the most popular optic-flow algorithm in computer vision is due to Li *et al.* [80], [81] considering its high accuracy and low computational burden. However, there are several issues in this algorithm. First, the algorithm is derived under an assumption of constant brightness, a requirement that is not quite relevant in practice due to noise as well as shadowing effects that can vary the intensity of the light. Although there are several approaches in the literature to overcome this problem, each has its own potential drawbacks, making this a fertile research area. For instance, Kima *et al.* modeled variations in brightness using multiplicative and additive parameters before simultaneously solving the motion and illumination fields. The solution turns out to be somewhat complicated considering the need to compute two additional parameters, and as a result, the system could misunderstand the variation in pixel density. Another potential solution is to preprocess information that is not sensitive to light variations, so that one does not need to modify the main body of the algorithm. Interested readers in this topic may refer to [13]–[15], [80], and [81].

The use of dual optic-flow sensors, orthogonally displaced to the optical axis, to sense the altitude and speed of a UAV simultaneously was reported by Chahl *et al.* [83]. This novel concept uses two downward looking sensors offset vertically by a known distance. As the optical flow obtained from a sensor nearer to the ground will provide faster angular movement compared with a more distant one, it is possible to derive the altitude to the ground by comparing the relative optic-flow magnitude of both the sensors. As a by-product, ground speed can also be measured. This enables both scale and speed to be calculated, provided the aircraft is close enough to the ground and flying fast enough for the disparity in optic-flow signal to be measurable above noise. This system may provide advantages over stereo vision as stereo vision requires two independently calibrated cameras while dual optical flow sensors need not be calibrated and do not require an overlapping region between the two visual fields. A drawback of this configuration is that the cameras require a substantial offset for altitudes above a few meters.

### C. 3D-Range Camera

Range cameras are sensors that output an image with depth (range) information in each pixel. A camera can provide illumination, depth, and even color information at each pixel. The latter type is known as the RGB-D camera, which stands for the red-green-blue-depth channels being provided. There are two main types of range cameras that have found widespread implementation in robotics: time-of-flight (ToF) cameras and structured light-based cameras as in [84]. A stereo camera can also be considered a form of range camera, which uses stereo disparity between two images taken from two locations to determine range at various points in the image, but the results may not be available in every pixel.

Considering the benefit of structured lighting-based depth cameras as low-cost sensors to calculate the depth map of the environments, Stowers *et al.* [85] employed a calibrated Kinect depth and image sensors to control the altitude of a quadcopter drone in a real-time application. The Microsoft Kinect employs the primesense chipset to form structured light, as well as a LightCoding technique to compute depth. It has an IR laser projector combined with a CMOS camera, and an RGB video camera with 640 × 480 pixel images at 30 Hz. A more comprehensive discussion relating to 3D-range cameras can be found in [84].

### D. Infrared Camera

One major advantage of IR vision is its reliability, since it does not depend on visible light, making it is also suitable for night vision. The effectiveness of the IR camera as a UAV sensor can be illustrated by the research conducted by Yakimenko *et al.* [86], who addressed the problem of determining the relative position of a UAV to a ship, given three hot points on the ship as reference points While initially the ship can be regarded as a single spot by the onboard IR camera, as it gets closer, more features can be determined giving more information on pose and relative distance. It is apparent that the IR vision system can be effectively used to detect a distant object, despite variations in visibility. However, this may become a less affordable avenue, since quality IR cameras are expensive.

## IV. State-of-the-Art Visual and Inertial Navigation Systems

Vision-based localization systems have been widely implemented for MAV indoor navigation. To make the system work autonomously, onboard processing is desirable. To support the development of high levels of autonomy, one will also need to equip the system with high-performance onboard computers to deal with various expensive image processing algorithms, without much intervention from the ground station. Comparing vision-based localization systems with other localization methods, one can argue that the accuracy is the main strength of the vision-based method. However, it is apparent that the cost and complexity may somehow limit their practical usefulness despite the steady decrease of the price of camera and computer hardware.

IMUs have been widely implemented to perform position estimation within outdoor and indoor environments. IMU systems alone, nonetheless, cannot be used as stand-alone localization systems, because they suffer from position and velocity drift over time unless correction is available from an external source, such as GPS or vision. Fusing IMU systems with other sensors offer several advantages, such as slowing the rate of update of the external system, leading to further extra benefits, e.g., reduced communication bandwidth, lower energy consumption at the system level, as well as scalability [87], [88].

In outdoor environments, an IMU can be used to supplement GPS systems (i.e., GPS/INS). Since we cannot rely on GPS or other external systems to become fully autonomous, vision holds the key. The main advantage of the visual/INS over GPS/INS is its self-contained nature (an ideal candidate for autonomous processing) in both indoor and outdoor environments. It is also free from signal masking problems as the system requires no external signal for positioning. In addition, the onboard cameras can be used to give an accurate estimate of the vertical distance and the movement of the UAV with respect to the ground [89]. Although this seems to be a computationally intensive solution (e.g., due to complex image processing algorithms), the speed and memory of the onboard computers have significantly improved, making this issue less relevant. In what follows, we will discuss some cutting-edge VI navigation systems in the literature. We briefly summarize our discussion in Table II.

Considering the importance of VI information fusion, Corke *et al.* [90], [91] highlighted the complementarity of visual/inertial sensors. While inertial sensors suffer from large measurement uncertainty at slow motion, it can perform better at higher speed [90]. On the other hand, cameras can perform accurate tracking at low velocities, and their performance is inversely proportional to the velocity due to the increased bandwidth that may complicate the image processing algorithms for real-time implementations. Yet another limitation of visual sensors is due to their inability to distinguish rotational from translational motion, since it requires the summation of six motion components. There is also a well-known near–far effect when using cameras that can lead to a difficulty to distinguish between a near object that moves slowly and a distant object that moves quickly.

Technically, the integration of VI sensors can be performed by integrating images obtained from a camera with pose and estimation from the inertial sensors. Fusing this information will simplify the reconstruction of the 3D environments. Besides, inertial sensors can provide important information relating to a scene structure in the form of vertical and horizontal references of orientation. Compared with information obtained by means of SfM algorithms, VI fusion can achieve similar outcomes, which substantially lower computational cost and reasonably higher robustness [92], [93].

Garratt *et al.* [14] designed a visual flight control system based on a snapshot image of the ground, as an anchor point. Subsequently, the translational movements of the helicopter as well as its velocity can be calculated by comparing the sequence of image frames with respect to the stored image before the data are fed into its onboard autopilot. The effectiveness of the algorithm was confirmed through simulation results and in flight tests for 2D and 3D snapshots. Nonetheless, one potential drawback of the snapshot hover algorithm is due to the need to bootstrap the algorithm by means of a measure of scale. For instance, the integration of vertical velocity to obtain the estimated initial height may limit the effectiveness of the algorithm, since MAV may not be able to move away more than 20% of the height away from the taken snapshot.

Meier *et al.* [94] developed an onboard computer vision system for autonomous flight of a PIXHAWK quadcopter in indoor environments, supported with a cutting-edge augmented reality marker, namely, ARToolkit markerboard. It is an instrument that can give the position of its marker in the camera coordinate system, with a relatively short delay within (5–10 ms) [94]. Markers act as artificial target features, which are organized in an array. To perform a localization task, the positions of the markers are encoded in a global world map, so that they can provide six-DOF orientations with respect to the global reference map. The system allows a tight integration of IMU measurements into a computer vision framework using a KF. The system also employs a stereo obstacle detection module to collect depth information. Considering the

TABLE II
STATE-OF-THE-ART APPLICATIONS OF VI NAVIGATION SYSTEMS: A COMPARATIVE STUDY

| No | Platfoms | Purpose | Navigation Systems | Pros | Cons |
|---|---|---|---|---|---|
| 1 | H-610 helicopter model [15] | To develop an autonomous hovering system by means of visual-inertial navigation | Optical flow - inertial | Low-cost, autonomous | Inherently noisy, need scale or velocity from another sensor for the optic-flow measurement to be useful. |
| 2 | PIXHAWK Quadcopter supported by ARToolkit markerboard [95] | To introduce a new autonomous quadcopter platform, known as, the Pixhawk MAV | Visual-inertial system | Fully Autonomous - no need for a constant support from the ground station, precise time-stamping, leading to an optimum data fusion in the absence of delay | Computationally intensive. |
| 3 | Self-customized indoor quadcopter [88] | To realize a robust indoor UAV system with real-time processing capability solely onboard | Visual-inertial odometry: inertial measurement units (IMU), a monocular camera and scanning laser rangefinder supported with Kalman Filters for data fusion among sensor | Simplicity, less computational burden, real-time processing capability solely onboard (self-sustained), less dependability with the ground station making it a more robust and reliable, flexibility or adaptability of the algorithms, weight reduction and power efficiency | Reasonably short flight duration. |
| 4 | X4-flyer Quadcopter [96] | Develop model-based tracking for indoor maneuvering | Visual-inertial odometry: a camera and an inertial measurement unit (IMU) supported with EKF and hierarchical control | Reasonably good performance, robustness to the noise produced by transmission interferences | Dependability on wireless link and ground station for image processing system may reduce its reliability, no ground truth to evaluate the precision of the position and velocity estimates. |
| 5 | MAV [7] | To develop an efficient navigation algorithm for a MAV | Visual-inertial odometry represented by a single camera and inertial measurement unit (IMU) | Cm accuracy, light-weight design, low power consumption, low implementation cost, scalable for 6 degree-of-freedom pose estimation with constant complexity, able to overcome the failure in the virtual SLAM (VSLAM) framework by online re-initialization, anti-interference and high-security system | Expensive in terms of image processing algorithm, a big latency between visual-inertial measurements. |
| 6 | Small-scaled helicopter Vario Benzin-Acrobatic 23cc [97] | To develop developed robust and non-linear fusion of visual/inertial data sensors | Visual-Inertial odometry comprising of accelerometer and gyroscope to estimate the position, velocity, and attitude for an unmanned helicopter | Produced a reasonably accurate performance demonstrated by the ability to simultaneously build 3D feature map in the absence of GPS signals and to bound the error of the vehicle position estimation | Computationally intensive. |
| 7 | Parrot AR Drone [98] | To develop a new algorithm for a small UAV to navigate in GPS-denied environments | Visual-Inertial odometry (fusion of monocular visual SLAM and IMU) | They demonstrated the observability of all motion patterns including the absolute scale; unlike other algorithms employing other visual-inertial fusion methods, that assumed zero accelerometer bias or required non-zero acceleration to guarantee the observability of the scale | Off-line process due to intensive computational process. |
| 8 | Simultaneous localisation and mapping (SLAM)-based Quadcopter [99] - autonomous approach | To preserve a large degree of flexibility of a small quadcopter to navigate indoor | Multi-level SLAM using propagation time of laser for indoor mapping equipped with inertial sensors as well as particle and Kalman filters | High-degree of flexibility for small indoor UAV | Computationally expensive system, the computational cost per frame is $\mathcal{O}(n)^3$ [35]. Thus, this may not be unsuitable for systems with limited computational payload. Also, dependability on ground station processing unit may lead to reliability issues. |
| 9 | Quadcopter MAV [100] | Robust and autonomous navigation system (i.e. ability to cope with processing failures for one pair of camera and error in pose estimation) | Four cameras in two stereo configurations (one pair of camera faces downward, while another pair faces forward as an input for a reduced stereo SLAM system) | Substantial increase of accuracy and robustness due to the complement nature of the configuration of the cameras, improved safety and reliability by adding more redundancy | Unsuitable for low lighting, foggy or insufficient texture environments, computationally intensive process. |
| 10 | AR.Drone Quadcopter [77] | To develop an open source software for a low-cost quadcopter while addressing the limitation of monocular cameras | Monocular camera, key-frame based visual SLAM | low-cost system, reasonably accurate | limitation due to the nature of PTAM, namely, the position of the camera is only up-to-scale in regards to the PTAM frame, and accordingly the system needs a scale estimation to derive the actual scale of the environments, dependency with respect to the field-of-view (FoV) of the camera. |

imperfection in vision matching process, the authors develop an outlier removal algorithm by comparing redundant data in roll and pitch estimates obtained from the IMU with respect to one obtained from visual localization. The authors performed real-time flight tests to highlight the benefits of the proposed system. The system can demonstrate precise time stamping of IMU-vision synchronization, allowing optimum fusion of VI information in the absence of delays. However, the system only works in cooperative environments.

Considering the flight of indoor MAVs, Wang *et al.* [87] develop a navigation system based on the integration of an IMU and a downward looking monocular camera as a means of estimating the body axis of the UAV (since optical flow estimates velocity accurately when the camera is orthogonal to the axis), and a scanning LRF supported with KF for optimum data fusion among sensors. The role of the KF is to perform optimal data fusion between the IMU and vision sensors in order to determine the estimated position required by its onboard control system. The system employs a scanning LRF for path planning. The main advantage of this algorithm is its low computational power, since the algorithm is relatively simple, so that it can be run autonomously by an onboard autopilot.

Weiss *et al.* [7] proposed an efficient navigation algorithm for an MAV, which relies on the combination of a single camera and IMU, which is capable of running onboard and in real time. Owing to the EKF framework for processing IMU data, the algorithm focuses on the speed estimation.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



The module can be used for self-calibration of the sensors in real time and is employed during initialization. The main advantage of this hardware configuration is its lightweight design, low-power consumption, and low implementation cost. The algorithm is also scalable for six-DOF pose estimation with constant complexity as evidenced by simulation and real experiments. Moreover, it presents an improvement on the existing monocular framework for drift-free position control. It can also overcome the failure in the virtual SLAM framework by online reinitialization using their inertial-optical flow approach for a short-term position hold.

Cheviron *et al.* [96] developed robust and nonlinear fusion of visual/inertial data sensors comprising of accelerometers and gyroscopes to estimate the position, velocity, and attitude for an unmanned helicopter (i.e., three-DMG Microstrain IMU and a Philips Web-cam mounted on a small-scaled helicopter Vario Benzin-Acrobatic 23cc) based on the implementation of coupled nonlinear observer. The algorithm also overcomes the strong bias in low-cost inertial measurement systems. While the first observer aims to estimate the orientation matrix and bias on the angular velocity observations, the second counterparts determine the coordinate position and the linear velocity on top of bias terms on the accelerometer observations. The convergence of the estimation is achieved by means of adaptive control and backstepping analysis. The authors investigated the ability of the UAV to hover over a visual target in black and white colors, and the experiment was conducted, such that the target was always within the view of the camera that was statically calibrated before the flight. The algorithm produced a reasonably accurate performance demonstrated by the ability to simultaneously build a 3D feature map in the absence of GPS signals and to bind the error of the vehicle position estimation, despite significant error caused by initial conditions and initial sensor drift and some periodic lacks of visual measurement.

Abeywardena *et al.* [97] developed a new algorithm for fusing monocular visual SLAM and IMU for a small UAV to navigate in GPS-denied environments. The authors employed a single camera to measure the motion of the vehicles and to detect building landmarks. They implemented the EKF-based SLAM to estimate the position of the UAV. The authors employed a dual axis accelerometer, mounted in parallel to the propeller of the vehicle. This way, they demonstrated the observability of all motion patterns, including the absolute scale, unlike other algorithms employing VI fusion methods that assumed zero accelerometer bias or required nonzero acceleration to guarantee the observability of the scale.

Schauwecker and Zell [99] presented a new indoor navigation system for a quadcopter MAV by means of two stereo cameras to overcome the limitation of both monocular and stereo cameras. The limitation of the monocular camera is that the system only allows the estimation of the position of MAV with respect to an unknown and unobservable scaling factor. This problem can be addressed with stereo vision to solve the scale problem. Mounting stereo cameras in forward-facing direction can provide a large FoV to enable the detection of obstacles in flying direction. However, fast yaw rotation can limit the effectiveness of forward-facing cameras, e.g., due to motion blur or large image movements. A downward-facing camera, nevertheless, has a better capability in dealing with yaw rotation, and could be used to measure pitch, roll, and altitude. Nonetheless, the result is poor for fast horizontal movements flight and for one in low ground proximity.

To address this issue, Schauwecker and Zell [99] presented a quadcopter MAV with two pairs of stereo cameras to split the task equally for both forward-facing and downward-facing cameras. The authors successfully demonstrated that the accuracy and robustness of the self-localization system can be substantially improved by adding an additional downward-facing camera, making their work to be the first in the literature to employ a stereo matching with more than two cameras onboard of an MAV. It should be noted that both forward looking and downward-facing cameras are complimentary to each other. While a pair of downward looking camera is intended for ground plane detection and tracking, a pair of forward looking counterparts is designated as an input for a stereo SLAM system. Although the nature of the system is more computationally intensive, it is also still fully autonomous, since all processing, including sparse stereo matching, can be performed onboard, in real time, at high processing rates. Furthermore, the system is able to recover from pose estimation error and is sufficiently robust to overcome processing failures for one pair of cameras.

## V. Discussion

In this section, we will discuss some potential research challenges that need to be considered when designing robust and autonomous visual/INSs for MAVs. Accordingly, there are also some potential research opportunities that may arise from them.

### *A. Characteristics of Missions: Indoor and Outdoor*

The topological condition of indoor flying environments, compared with outdoor counterparts, is more challenging due to limited space, narrow flight terrain, limited height, as well as the presence of numerous obstacles. This will lead to a more specific design requirement, such as limited speed range, that is, typical walking speed of human in the range of 0–0.5 m/s compared with outdoor UAVs, which fall far above this range, e.g., 3–45 m/s [3]. Thus, to successfully accomplish indoor mission planning, careful consideration in designing UAV's physical and aerodynamic structures is also of paramount importance. This will create a research opportunity in the selection of appropriate sensors and actuators to meet the demand of indoor flight environments [100].

When it comes to tracking accuracy, the demand shall vary from case to case. While centimeter resolution may not be necessary for search-and-rescue scenarios, whose mission is to locate the coordinate position of people at room level [101], in many other applications, e.g., for building inspections and mapping, it would be quite essential [9], [102]. Although, there have been significant attempts to mitigate the adverse impacts of multipath fading for indoor applications, the researchers still end up with something less favorable, that is, having complicated and computationally intensive algorithms [34]. This fact is indeed undesirable, since most of the indoor UAVs are subject to limited mechanical and electrical payloads [101], [102]. Thus, having a complex algorithm could also limit the practical usefulness of the system. In addition, variations in visibility (e.g., fog) can lead to challenging circumstances, particularly, when it comes to the accuracy of the visual systems to detect or track an object of interest. Disparity in weather conditions, time of day, changes in altitude and whether the flight is indoors or in the shade can have a dramatic effect on the brightness level presented to camera. Since time is critical in target tracking, the ability of the navigation systems to quickly adapt various brightness levels is essential.

### *B. Reliability Against Limited Communication Range and Sensors Failures*

System reliability in the face of limited communication range is also a potential research avenue as one cannot guarantee a continuous stable broadband radio link in all environments. For typical deployment of indoor UAVs in large buildings, their limited communication range can sometimes become problematic [103]. It is often deteriorated by the use of separate processing computers for the onboard autopilot and the ground station (normally used to perform computationally intensive image processing tasks before sending the results back to the onboard autopilots). Solving this problem using a set of relays of communication will present new research challenges, such as optimum relays allocation problems [103]. Network constraints imposed, such as bandwidth limitation, also become a very important consideration as no one can afford to have unlimited bandwidth [53], [103]–[105].

### *C. Autonomous and Intelligent Versus Distributed (Infrastructure-Assisted) Approaches*

While most people tend to make the system autonomous, resulting in intelligent onboard processing, some also consider a distributed (infrastructure-assisted) approach for its reliability. Considering autonomous and intelligent systems, self-contained positioning systems, such as VI odometry (i.e., hybrid cameras and IMU), are ideal candidates to support the mission. Leveraging the benefits of hybrid positioning systems, one possible research avenue worth considering is the development of optimal fusion of data of visual/inertial sensors [7], [87], [95], [98], [106]. Given the steady increase in the speed of computers as well as advancements in sensor technology, we can expect to witness more and more results to come in this area.

However, depending on the objectives of the missions, one may also consider the development of distributed (infrastructure)-based approach, such as by means of wireless ZigBee networks in [88], RFID in [107], and numerous other approaches [53], [108], [109]. Thus, creating optimum fusion of data among multiple heterogeneous sensors for distributed navigation of indoor MAV is also considered a fertile research area, e.g., IMU with distributed wireless sensor networks [110]. Although inertial sensor suffers from drift and thus its accuracy considerably decreases against time, it can be implemented within a short period of time to enhance the accuracy of the system. This, in turn, will also give the opportunity to relax the sampling rate of the system and operate the measurements with lower bit overheads to reduce the traffic of communications, and to conserve more energy.

### *D. Safety and Integrity Issues*

The safety of UAV operations is enforced by rigid regulations issued by a certain authority [111]. For instance, in the U.S., the Federal Aviation Administration regulates and oversees all aspects of American civil aviation to promote safety, while the Civil Aviation Safety Authority (CASA) is the primary body for the maintenance, enhancement, and promotion of the safety civil aviation in Australia, including civilian UAV operation.

Visual-inertial is in the early stages of development and will one day need to be certified in the future in the same way that GPS has been certified as a navigation aid for aviation. Vision might fit into the overall safety paradigm for aviation, e.g., as a redundant backup to GPS [89]. Although visual systems are more brittle than technologies, such as GPS, they can provide data that GPS cannot, such as range to obstacles and relative terrain height in dynamic cluttered environments [89], [112], [113].

Given the vital role of aerial robots, it is expected that in the future, there will be a group of UAVs operating side-by-side with manned aircraft to accomplish their challenging missions [111]. This signifies the importance of the safety and integrity issues in the development of both manned and UAVs. However, MAVs have low kinetic energy and do not need the same safety regulation as larger UAVs and manned operations. The main risk is collision and resulting injury and property damage, but obviously something the size of an MAV cannot do much damage. New regulations from CASA have been considered and do not legislate for vehicles under a certain size (e.g., 2 kg in Australia). For vehicles in the MAV range, the risk of air-to-air collision is akin to a bird strike on a general aviation or commercial flight. Finally, UAVs employ no onboard aircrew; hence there is no hazard to operating personnel.

## VI. Conclusion and Future Trend

The developments of low-cost MAVs supported by recent advancements in visual/inertial navigation have greatly assisted many aspects of modern people's lives, i.e., safety and productivity, particularly in the environments where people face greater risks and challenges, e.g., building inspections and monitoring, battlefields, indoor farming, underground mining, as well as search-and-rescue scenario in disaster areas [101].

It is also interesting to note that many current systems are much akin to the principle of biological systems in nature, such as the navigation systems of insects and birds, which depend on the spatial, temporal, and spectral distribution of light in their navigation [114]. For instance, a review on the memory use in insect visual navigation can be found in [115], while a study on biomimetic visual sensing and flight control is presented in [116], and the development of three biomimetic flight control sensors is depicted in [114].

We envisage that robust, accurate, and yet inexpensive visual/inertial MAV navigation systems specifically designed to navigate in cluttered environments will gradually become an integral part in modern people's lives. This will result in increased safety and productivity as well as comfort. We believe that hybrid systems, such as visual/inertial systems, will predominate future markets. This works mainly because of mutual complementarity to cover the deficiencies of other sensors. In addition, as the speed of computers and the technology of MEMS sensors progressively advance, we can achieve highly accurate and robust positioning systems supported with sophisticated computing algorithms, such as particle filters, which once were considered computationally expensive and thus hardly implemented for real-time applications.

## REFERENCES


[1] J. M. McMichael and M. S. Francis, "Micro air vehicles—Toward a new dimension in flight," Defense Adv. Res. Projects Agency, Arlington, VA, USA, Tech. Rep. 8/7/97, 1997.

[2] F. Santoso, M. Liu, and G. Egan, "Optimal control linear quadratic synthesis for an unconventional aircraft," in *Proc. 12th Austral. Int. Aerosp. Congr. (AIAC)*, Mar. 2006.

[3] F. Santoso, M. Liu, and G. Egan, "$H_2$ and $H_\infty$ robust autopilot synthesis for longitudinal flight of a special unmanned aerial vehicle: A comparative study," *IET Control Theory Appl.*, vol. 2, no. 7, pp. 583–594, 2008.

[4] F. Santoso, M. Liu, and G. K. Egan, "Robust $\mu$-synthesis loop shaping for altitude flight dynamics of a flying-wing airframe," *J. Intell. Robot. Syst.*, vol. 79, no. 2, pp. 259–273, Aug. 2015.

[5] M. Liu, G. K. Egan, and F. Santoso, "Modeling, autopilot design, and field tuning of a UAV with minimum control surfaces," *IEEE Trans. Control Syst. Technol.*, vol. 23, no. 6, pp. 2353–2360, Nov. 2015.

[6] A. Malaver, N. Motta, P. Corke, and F. Gonzalez, "Development and integration of a solar powered unmanned aerial vehicle and a wireless sensor network to monitor greenhouse gases," *Sensors*, vol. 15, no. 2, pp. 4072–4096, 2015.

[7] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time onboard visual-inertial state estimation and self-calibration of MAVs in unknown environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, Saint Paul, MN, USA, May 2012, pp. 957–964.

[8] S. Lin, M. A. Garratt, and A. J. Lambert, "Real-time 6DoF deck pose estimation and target tracking for landing an UAV in a cluttered shipboard environment using on-board vision," in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Beijing, China, Aug. 2015, pp. 474–481.

[9] C. Eschmann, C.-M. Kuo, C.-H. Kuo, and C. Boller, "Unmanned aircraft systems for remote building inspection and monitoring," in *Proc. 6th Eur. Workshop Struct. Health Monitor.*, Dresden, Germany, 2012, pp. 1–8.

[10] C. Zhang and J. M. Kovacs, "The application of small unmanned aerial systems for precision agriculture: A review," *Precis. Agricult.*, vol. 13, no. 6, pp. 693–712, Dec. 2012.

[11] The Australian Government. (2015). *Civil Aviation Safety Authority, UAV Camera Inside Outback Australia*. [Online]. Available: http://uavcamera.com.au/

[12] M. A. Garratt and J. S. Chahl, "Vision-based terrain following for an unmanned rotorcraft," *J. Field Robot.*, vol. 25, nos. 4–5, pp. 284–301, Apr./May 2008.

[13] M. A. Garratt and A. Cheung, "Obstacle avoidance in cluttered environments using optic flow," in *Proc. Australasian Conf. Robot. Autom. (ACRA)*, Sydney, NSW, Australia, Dec. 2009, pp. 1–9.

[14] M. A. Garratt, A. J. Lambert, and H. Teimoori, "Design of a 3D snapshot based visual flight control system using a single camera in hover," *Auto. Robots*, vol. 34, no. 1, pp. 19–34, Jan. 2013.

[15] M. A. Garratt and J. S. Chahl, "An optic flow damped hover controller for an autonomous helicopter," in *Proc. 22nd Int. UAV Syst. Conf.*, Apr. 2007, pp. 16–18.

[16] W. E. Green, P. Y. Oh, and G. Barrows, "Flying insect inspired vision for autonomous aerial robot maneuvers in near-earth environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, New Orleans, LA, USA, Apr./May 2004, pp. 2347–2352.

[17] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA, USA: Artech House, 2004.

[18] L. Smith and V. Aitken, "The auxiliary extended and auxiliary unscented Kalman particle filters," in *Proc. Can. Conf. Elect. Comput. Eng. (CCECE)*, Apr. 2007, pp. 1626–1630.

[19] G. Evensen, *Data Assimilation: The Ensemble Kalman Filter*. Berlin, Germany: Springer-Verlag, 2007.

[20] A. Budhiraja, L. Chen, and C. Lee, "A survey of numerical methods for nonlinear filtering problems," *Phys. D, Nonlinear Phenomena*, vol. 230, nos. 1–2, pp. 27–36, Jun. 2007.

[21] K. Y. K. Leung, F. Inostroza, and M. Adams, "An improved weighting strategy for Rao–Blackwellized probability hypothesis density simultaneous localization and mapping," in *Proc. Int. Conf. Control, Autom., Inf. Sci.*, Nov. 2013, pp. 103–110.

[22] F. Daum, "Nonlinear filters: Beyond the Kalman filter," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 20, no. 8, pp. 57–69, Aug. 2005.

[23] G. G. Rigatos, "Distributed filtering over sensor networks for autonomous navigation of UAVs," *Intell. Service Robot.*, vol. 5, no. 3, pp. 179–198, Jul. 2012.

[24] D. J. Allerton and H. Jia, "A review of multisensor fusion methodologies for aircraft navigation systems," *J. Navigat.*, vol. 58, no. 3, pp. 405–417, 2005.

[25] P. Zhan, D. W. Casbeer, and A. L. Swindlehurst, "A centralized control algorithm for target tracking with UAVs," in *Proc. 39th Asilomar Conf. Signals, Syst. Comput.*, pp. 1148–1152, Oct./Nov. 2005.

[26] Y. Zhai, M. B. Yeary, J. P. Havlicek, and G. Fan, "A new centralized sensor fusion-tracking methodology based on particle filtering for power-aware systems," *IEEE Trans. Instrum. Meas.*, vol. 57, no. 10, pp. 2377–2387, Oct. 2008.

[27] N. A. Carlson, "Federated filter for distributed navigation and tracking applications," in *Proc. 28th Annu. Meeting CIGTF 21st Guid. Test Symp.*, 2002, pp. 340–353.

[28] M. E. Liggins, II, C.-Y. Chong, I. Kadar, M. G. Alford, V. Vannicola, and S. Thomopoulos, "Distributed fusion architectures and algorithms for target tracking," *Proc. IEEE*, vol. 85, no. 1, pp. 95–107, Jan. 1997.

[29] D. Simon, *Optimal State Estimation: Kalman, $H_\infty$, and Nonlinear Approaches*. New York, NY, USA: Wiley, 2006.

[30] J. Kim, L. L. Ong, E. Nettleton, and S. Sukkarieh, "Decentralized approach to UAV navigation: Without the use of GPS and preloaded maps," *Proc. Inst. Mech. Eng. G, Aerosp. Eng.*, vol. 218, no. 6, pp. 399–416, Dec. 2004.

[31] S. Sukkarieh, E. Nettleton, J.-H. Kim, M. Ridley, A. Goktogan, and H. Durrant-Whyte, "The ANSER project: Data fusion across multiple uninhabited air vehicles," *Int. J. Robot. Res.*, vol. 22, nos. 7–8, pp. 505–538, Jul./Aug. 2003.

[32] K. J. Åström and B. Wittenmark, *Adaptive Control*. Reading, MA, USA: Addison-Wesley, 1995.

[33] G. Welch and G. Bishop, "An introduction to the Kalman filter," Dept. Comput. Sci., Univ. North Carolina Chapel Hill, Chapel Hill, NC, USA, Tech. Rep. TR 95-041, 1995.

[34] H. Strasdat, J. M. M. Montiel, and A. J. Davison, "Real-time monocular SLAM: Why filter?" in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2010, pp. 2657–2664.

[35] O. Straka, J. Duník, and M. Šimandl, "Gaussian sum unscented Kalman filter with adaptive scaling parameters," in *Proc. 14th Int. Conf. Inf. Fusion*, Chicago, IL, USA, Jul. 2011, pp. 1–8.

[36] C. Hajiyeva and H. E. Sokenb, "Robust Adaptive Kalman Filter for estimation of UAV dynamics in the presence of sensor/actuator faults," *Aerosp. Sci. Technol.*, vol. 28, no. 1, pp. 376–383, Jul. 2013.

[37] K. Xiong, H. Zhang, and L. Liu, "Adaptive robust extended Kalman filter for nonlinear stochastic systems," *IET Control Theory Appl.*, vol. 2, no. 3, pp. 239–250, 2008.

[38] G. Huang, M. Kaess, and J. J. Leonard, "Towards consistent visual-inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2014, pp. 4926–4933.

[39] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint Kalman filter for vision-aided inertial navigation," in *Proc. IEEE Int. Conf. Robot. Autom.*, Apr. 2007, pp. 3565–3572.

[40] L. R. G. Carrillo, A. E. D. López, R. Lozano, and C. Pégard, *Quad Rotorcraft Control: Vision-Based Hovering and Navigation*. London, U.K.: Springer-Verlag, 2013.

[41] M. Zeitz, "The extended Luenberger observer for nonlinear systems," *Syst. Control Lett.*, vol. 9, no. 2, pp. 149–156, 1987.

[42] A. U. Awan, J. Park, and H. J. Kim, "Thrust estimation of quadrotor UAV using adaptive observer," in *Proc. 11th Int. Conf. Control, Autom. Syst.*, Gyeonggi-do, South Korea, Oct. 2011, pp. 131–136.

[43] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
[44] D. Bagnell and B. Wagenknecht, "SLAM, fast SLAM, and Rao-Blackwellization," *Statist. Techn. Robot.*, 2009.
[45] C. Kim, R. Sakthivel, and W. K. Chung, "Unscented FastSLAM: A robust and efficient solution to the SLAM problem," *IEEE Trans. Robot.*, vol. 24, no. 4, pp. 808–820, Aug. 2008.
[46] A. Soto, "Self adaptive particle filter," in *Proc. 19th Int. Joint Conf. Artif. Intell.*, 2005, pp. 1398–1403.
[47] Z. Xiaowei, L. Hong, and S. Xiaohong, "Object tracking with an evolutionary particle filter based on self-adaptive multi-features fusion," *Int. J. Adv. Robot. Syst.*, vol. 10, no. 61, pp. 1–11, Jan. 2013.
[48] N. Abdelkrim, N. Aouf, A. Tsourdos, and B. White, "Robust nonlinear filtering for INS/GPS UAV localization," in *Proc. 16th Medit. Conf. Control Autom.*, Ajaccio, France, Jun. 2008, pp. 695–702.
[49] J. Seo, M.-J. Yu, C. G. Park, and J. G. Lee, "An extended robust H infinity filter for nonlinear uncertain systems with constraints," in *Proc. 44th IEEE Conf. Decision Control, Eur. Control Conf.*, Dec. 2005, pp. 1935–1940.
[50] Q. C. Wang, J. Li, M. X. Zhang, and C. H. Yang, "H-infinity filter based particle filter for maneuvering target tracking," *Prog. Electromagn. Res. B*, vol. 30, pp. 103–116, Jun. 2011.
[51] S. Chen, Y. Li, and N. M. Kwok, "Active vision in robotic systems: A survey of recent developments," *Int. J. Robot. Res.*, vol. 30, no. 11, pp. 1343–1377, 2011.
[52] E. A. Wan and R. Van Der Merwe, "The unscented Kalman filter for nonlinear estimation," in *Proc. IEEE Adapt. Syst. Signal Process., Commun., Control Symp.*, Oct. 2000, pp. 153–158.
[53] S. C. Lee, W. R. Lee, and K. H. You, "TDoA based UAV localization using dual-EKF algorithm," *Int. J. Control Autom.*, vol. 2, no. 4, pp. 47–54, 2009.
[54] J. Uhlmann. (Jan. 2015). *First-Hand: The Unscented Transform*, accessed on Jan. 1, 2016. [Online]. Available: http://ethw.org/First-Hand:The_Unscented_Transform
[55] F. Danescu, S. Oniga, S. Nedevschi, and M.-M. Meinecke, "Tracking multiple objects using particle filters and digital elevation maps," in *Proc. IEEE Intell. Vehicles Symp. (IV)*, Xi'an, China, Jun. 2009, pp. 88–93.
[56] F. Gustafsson *et al.*, "Particle filters for positioning, navigation, and tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 425–437, Feb. 2002.
[57] F. Gustafsson and F. Gunnarsson, "Mobile positioning using wireless networks: Possibilities and fundamental limitations based on available wireless network measurements," *IEEE Signal Process. Mag.*, vol. 22, no. 4, pp. 41–53, Jul. 2005.
[58] F. Santoso, "A new framework for rapid wireless tracking verifications based on optimized trajectories in received signal strength measurements," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 45, no. 11, pp. 1424–1436, Nov. 2015.
[59] F. Santoso and R. Malaney, "Tracking-based wireless intrusion detection for vehicular networks," in *Proc. IEEE Veh. Technol. Conf. (VTC)*, San Francisco, CA, USA, Sep. 2011, pp. 1–5.
[60] B. D. Gouveia, D. Portugal, D. C. Silva, and L. Marques, "Computation sharing in distributed robotic systems: A case study on SLAM," *IEEE Trans. Autom. Sci. Eng.*, vol. 12, no. 2, pp. 410–422, Apr. 2015.
[61] A. J. Haug, *Bayesian Estimation and Tracking: A Practical Guide*. New York, NY, USA: Wiley, Jun. 2012.
[62] A. Correia, J. Matias, P. Mestre, and C. Serodio, "Derivative-free nonlinear optimization filter simplex," *Int. J. Appl. Math. Comput. Sci.*, vol. 20, no. 4, pp. 679–688, 2010.
[63] H.-P. Chiu, S. Williams, F. Dellaert, S. Samarasekera, and R. Kumar, "Robust vision-aided navigation using sliding-window factor graphs," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2013, pp. 46–53.
[64] J. B. Moore, "Fixed-lag smoothing results for linear dynamical systems," *Austral. Telecommun. Res.*, vol. 7, no. 2, pp. 16–21, 1973.
[65] E. D. Nerurkar, K. J. Wu, and S. I. Roumeliotis, "C-KLAM: Constrained keyframe-based localization and mapping," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2014, pp. 3638–3643.
[66] T.-C. Dong-Si and A. I. Mourikis, "Motion tracking with fixed-lag smoothing: Algorithm and consistency analysis," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2011, pp. 5655–5662.
[67] R. Mottaghi, M. Kaess, A. Ranganathan, R. Roberts, and F. Dellaert, "Place recognition-based fixed-lag smoothing for environments with unreliable GPS," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, May 2008, pp. 1862–1867.
[68] K. Konolige, M. Agrawal, and J. Sola, "Large-scale visual odometry for rough terrain," in *Robotics Research*. Berlin, Germany: Springer-Verlag, 2011, pp. 201–212.
[69] E. S. Jones and S. Soatto, "Visual-inertial navigation, mapping and localization: A scalable real-time causal approach," *Int. J. Robot. Res.*, vol. 30, no. 4, pp. 407–430, Apr. 2011.
[70] A. Martinelli, "State estimation based on the concept of continuous symmetry and observability analysis: The case of calibration," *IEEE Trans. Robot.*, vol. 27, no. 2, pp. 239–255, Apr. 2011.
[71] A. Martinelli, "Observability properties and deterministic algorithms in visual-inertial structure from motion," *Found. Trends Robot.*, vol. 3, no. 3, pp. 139–209, Dec. 2013.
[72] A. Martinelli, "Closed-form solution of visual-inertial structure from motion," *Int. J. Comput. Vis.*, vol. 106, pp. 138–152, Jan. 2014.
[73] J. Engel, T. Schöps, and D. Cremers, "LSD-SLAM: Large-scale direct monocular SLAM," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2014, pp. 834–849.
[74] R. A. Newcombe, S. J. Lovegrove, and A. J. Davison, "DTAM: Dense tracking and mapping in real-time," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Nov. 2011, pp. 2320–2327.
[75] M. Irani and P. Anandan, "Robust multi-sensor image alignment," in *Proc. Int. Conf. Comput. Vis. (ICCV)*, Jan. 1998, pp. 959–966.
[76] I. Sa, H. He, V. Huynh, and P. Corke, "Monocular vision based autonomous navigation for a cost-effective MAV in GPS-denied environments," in *Proc. IEEE/ASME Int. Conf. Adv. Intell. Mechatronics (AIM)*, Wollongong, NSW, Australia, Jul. 2013, pp. 1355–1360.
[77] D. Scaramuzza *et al.*, "Vision-controlled micro flying robots: From system design to autonomous navigation and mapping in GPS-denied environments," *IEEE Robot. Autom. Mag.*, vol. 21, no. 3, pp. 26–40, Sep. 2014.
[78] H. Lim, J. Lim, and H. J. Kim, "Real-time 6-DOF monocular visual SLAM in a large-scale environment," in *Proc. Int. Conf. Robot. Autom. (ICRA)*, May/Jun. 2014, pp. 1532–1539.
[79] P. Gohl *et al.*, "Towards autonomous mine inspection," in *Proc. 3rd Int. Conf. Appl. Robot. Power Ind. (CARPI)*, Oct. 2014, pp. 1–6.
[80] P. Li, M. Garrat, and A. Lambert, "Monocular snapshot-based sensing and control of hover, takeoff, and landing for a low-cost quadrotor," *J. Field Robot.*, vol. 32, no. 7, pp. 984–1003, Oct. 2015.
[81] P. Li, M. Garratt, and A. Lambert, "A homography-based visual inertial fusion method for robust sensing of a micro aerial vehicle," in *Proc. IEEE Int. Conf. Mechatronics Autom.*, Beijing, China, Aug. 2015, pp. 488–493.
[82] Y.-H. Kim, A. M. Martínez, and A. C. Kak, "Robust motion estimation under varying illumination," *Image Vis. Comput.*, vol. 23, no. 4, pp. 365–375, Apr. 2005.
[83] J. Chahl, K. Rosser, and A. Mizutani, "Axially displaced optical flow sensors for feasuring and controlling the landing height of a UAV," in *Proc. 15th Austral. Int. Aerosp. Congr. (AIAC)*, 2013, pp. 221–227.
[84] F. Santoso, M. A. Garratt, M. R. Pickering, and M. Asikuzzaman, "3D mapping for visualization of rigid structures: A review and comparative study," *IEEE Sensors J.*, vol. 16, no. 6, pp. 1484–1507, Mar. 2016.
[85] J. Stowers, M. Hayes, and A. Bainbridge-Smith, "Altitude control of a quadrotor helicopter using depth map from microsoft Kinect sensor," in *Proc. IEEE Int. Conf. Mechatronics*, Istanbul, Turkey, Apr. 2011, pp. 358–362.
[86] O. A. Yakimenko, I. I. Kaminer, W. J. Lentz, and P. A. Ghyzel, "Unmanned aircraft navigation for shipboard landing using infrared vision," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 38, no. 4, pp. 1181–1200, Oct. 2002.
[87] F. Wang, J. Cui, S. K. Phang, B. M. Chen, and T. H. Lee, "A mono-camera and scanning laser range finder based UAV indoor navigation system," in *Proc. Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, Atlanta, GA, USA, May 2013, pp. 694–701.
[88] L. Yu, Q. Fei, and Q. Geng, "Combining Zigbee and inertial sensors for quadrotor UAV indoor localization," in *Proc. 10th IEEE Int. Conf. Control Autom. (ICCA)*, Hangzhou, China, Jun. 2013, pp. 1912–1916.
[89] J. Wang, M. Garratt, A. Lambert, J. J. Wang, S. Han, and D. Sinclair, "Integration of GPS/INS/vision sensors to navigate unmanned aerial vehicles," in *Proc. Int. Arch. Photogram., Remote Sens. Spatial Inf. Sci.*, vol. 38. Beijing, China, 2008, pp. 963–970.
[90] P. Corke, J. Lobo, and J. Dias, "An introduction to inertial and visual sensing," *Int. J. Robot. Res.*, vol. 26, no. 6, pp. 519–535, Jun. 2007.
[91] P. Corke, "An inertial and visual sensing system for a small autonomous helicopter," *J. Robot. Syst.*, vol. 21, no. 2, pp. 43–51, Feb. 2004.

[92] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, U.K.: Cambridge Univ. Press, 2011.
[93] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment—A modern synthesis," in *Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman, and R. Szeliski, Eds. Heidelberg, Germany: Springer-Verlag, Apr. 2002.
[94] L. Meier, P. Tanskanen, L. Heng, G. H. Lee, F. Fraundorfer, and M. Pollefeys, "PIXHAWK: A micro aerial vehicle design for autonomous flight using onboard computer vision," *Auto. Robots*, vol. 33, nos. 1–2, pp. 21–39, Aug. 2012.
[95] C. Teulière, L. Eck, E. Marchand, and N. Guénard, "3D model-based tracking for UAV position control," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Taipei, Taiwan, Oct. 2010, pp. 1084–1089.
[96] T. Cheviron, T. Hamel, R. Mahony, and G. Baldwin, "Robust nonlinear fusion of inertial and visual data for position, velocity and attitude estimation of UAV," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Rome, Italy, Apr. 2007, pp. 2010–2016.
[97] D. Abeywardena, Z. Wang, S. Kodagoda, and G. Dissanayake, "Visual-inertial fusion for quadrotor micro air vehicles with improved scale observability," in *Proc. IEEE Int. Conf. Robot. Autom. (ICRA)*, Karlsruhe, Germany, May 2013, pp. 3148–3153.
[98] S. Grzonka, G. Grisetti, and W. Burgard, "A fully autonomous indoor quadrotor," *IEEE Trans. Robot.*, vol. 28, no. 1, pp. 90–100, Feb. 2012.
[99] K. Schauwecker and A. Zell, "On-board dual-stereo-vision for autonomous quadrotor navigation," in *Proc. Int. Conf. Unmanned Aircraft Syst. (ICUAS)*, Atlanta, GA, USA, May 2013, pp. 333–342.
[100] A. Abbas, J. de Vicente, and E. Valero, "Aerodynamic technologies to improve aircraft performance," *Aerosp. Sci. Technol.*, vol. 28, no. 1, pp. 100–132, Jul. 2013.
[101] T. Tomic *et al.*, "Toward a fully autonomous UAV: Research platform for indoor and outdoor urban search and rescue," *IEEE Robot. Autom. Mag.*, vol. 19, no. 3, pp. 46–56, Sep. 2012.
[102] V. Kumar and N. Michael, "Opportunities and challenges with autonomous micro aerial vehicles," *Int. J. Robot. Res.*, vol. 31, no. 11, pp. 1279–1291, 2012.
[103] O. Burdakov, P. Doherty, K. Holmberg, J. Kvarnström, and P.-R. Olsson, "Positioning unmanned aerial vehicles as communication relays for surveillance tasks," in *Proc. Robot., Sci. Syst.*, Seattle, WA, USA, Jun. 2009, pp. 1–8.
[104] R. Kaune and A. Charlish, "Online optimization of sensor trajectories for localization using TDOA measurements," in *Proc. 6th Int. Conf. Inf. Fusion (FUSION)*, Istanbul, Turkey, Jul. 2013, pp. 484–491.
[105] F. Fletcher, B. Ristic, and D. Musicki, "Recursive estimation of emitter location using TDOA measurements from two UAVs," in *Proc. 10th Int. Conf. Inf. Fusion*, Quebec, QC, Canada, Jul. 2007, pp. 1–8.
[106] J. Rydell and E. Emilsson, "CHAMELEON: Visual-inertial indoor navigation," in *Proc. Position Location Navigat. Symp. (PLANS)*, Myrtle Beach, SC, USA, Apr. 2012, pp. 541–546.
[107] J. S. Choi, B. R. Son, H. K. Kang, and D. H. Lee, "Indoor localization of unmanned aerial vehicle based on passive UHF RFID systems," in *Proc. 9th Int. Conf. Ubiquitous Robots Ambient Intell. (URAI)*, Daejeon, South Korea, Nov. 2012, pp. 188–189.
[108] N. Kirchner and T. Furukawa, "Infrared localisation for indoor UAVs," in *Proc. 1st Int. Conf. Sens. Technol.*, Palmerston North, New Zealand, Nov. 2005, pp. 60–65.
[109] D. Liu and Y.-B. Li, "Research of indoor airship control system based on AP10," in *Proc. Int. Conf. E-Product E-Service E-Entertainment (ICEEE)*, Henan, China, Nov. 2010, pp. 1–3.
[110] F. Santoso, "Sub-optimal decentralised control algorithms for blanket and k-barrier coverage in autonomous robotic wireless sensor networks," *IET Commun.*, vol. 4, no. 17, pp. 2041–2057, Nov. 2010.
[111] R. Loh, Y. Bian, and T. Roe, "Safety requirements for unmanned aerial vehicles (UAV) in future civil airspace," in *Proc. IEEE/ION Position, Location, Navigat. Symp.*, San Diego, CA, USA, Apr. 24–27, 2006, pp. 1151–1163.
[112] M. H. Tehrani, M. A. Garratt, and S. G. Anavatti, "Integrated low-altitude horizon-based aircraft attitude estimation algorithms using UV-filtered panoramic images and optic flow," *IEEE Trans. Aerosp. Electron. Syst.*, to be published.
[113] P. Williams and M. Crump, "Intelligent landing system for landing UAVs at unsurveyed airfields," in *Proc. 28th Int. Congr. Aeronautical Sci.*, 2012, pp. 1–19.
[114] J. Chahl, "Three biomimetic flight control sensors," *Int. J. Intell. Unmanned Syst.*, vol. 2, no. 1, pp. 27–39, 2014.
[115] T. S. Collett and M. Collett, "Memory use in insect visual navigation," *Nature Rev. Neurosci.*, vol. 3, no. 7, pp. 542–552, Jul. 2002.
[116] G. L. Barrows, J. S. Chahl, and M. V. Srinivasan, "Biomimetic visual sensing and flight control," in *Proc. Bristol UAV Conf.*, 2002, pp. 1–15.

**Fendy Santoso** received the master's degree in electrical and computer systems engineering from Monash University, Melbourne, VIC, Australia, in 2007, supported by an Australian Development Scholarship from the Australian Government, and the Ph.D. degree in electrical engineering from the University of New South Wales (UNSW), Sydney, NSW, Australia, in 2012, under a University International Postgraduate Award Scholarship.

He has been a Post-Doctoral Researcher with the School of Engineering and Information Technology, UNSW, Australian Defence Force Academy, Canberra, ACT, Australia, since 2015. His current research interests include flight control systems and navigation of aerial robotics.

Dr. Santoso has been a Reviewer for many control and robotics journals, such as the IEEE ROBOTICS AND AUTOMATION LETTERS and the *International Journal of System, Control, and Communications*, and many international conferences.

**Matthew A. Garratt** received the Ph.D. degree in robotics from Australian National University (ANU), Canberra, ACT, Australia, in 2007.

He was an Aeronautical Engineer with the Royal Australian Navy, Canberra, for over a decade, before he was involved in the field of computer-aided engineering tools for two years. In 1999, he joined ANU, where he helped build and test an autonomous helicopter using biologically inspired vision for a Defense Advanced Research Projects Agency funded project. Since 2001, he has been with the School of Engineering and Information Technology, University of New South Wales, Canberra, as a Lecturer. His current research interests include sensing and control for autonomous systems.

**Sreenatha G. Anavatti** was with the Indian Institute of Technology Mumbai, India, as an Associate Professor. He moved to Australia in 1998. He is a Senior Lecturer with the School of Engineering and Information Technology, University of New South Wales, Canberra, ACT, Australia. He has supervised a number of Ph.D. scholars over the years. His current research interests include the application of artificial intelligence techniques, such as fuzzy and neural systems for unmanned aerial vehicles, underwater vehicles, and ground vehicles, identification and control of dynamic systems, navigation and path planning for autonomous vehicles, and evolutionary fuzzy systems and their applications to dynamic systems.