Sensor Modalities and Fusion for Robust Indoor Localisation

Michał Kozłowski*, Raúl Santos-Rodríguez, Robert J. Piechocki

Digital Health Engineering Group, University of Bristol, 1 Cathedral Square, Bristol, BS1 5DL

Abstract

The importance of accurate and efficient positioning and tracking is widely understood. However, there is a pressing lack of progress in the standardisation of methods, as well as generalised framework of their evaluation. The aim of this survey is to discuss the currently prevalent and emerging types of sensors used for location estimation. The intent of this review is to take account of this taxonomy and to provide a wider understanding of the current state-of-the-art. To that end, we outline various sensor modalities, as well as popular fusion and integration techniques, discussing how their combinations can help in various application settings. Firstly, we present the fundamental mechanics behind sensors employed by the localisation community. Furthermore we outline the formal theory behind prominent fusion methods and provide exhaustive implementation examples of each. Finally, we provide points for future discussion regarding localisation sensing, fusion and integration methods.

Keywords: Localisation, Efficiency, Fusion, Robustness

1. Introduction

Indoor localisation has been regularly cited as an important ambition of many fields in both, academia and industry. The use cases include pervasive health monitoring [13, 78], targeted advertising [9], factory vehicle tracking [71] and robotics [15, 90], amongst others. However, implementations of localisation methods and algorithms differ, depending on the need, deployment methods, available utilities, resources and sensors [54, 181].

At the heart of every implementation lies effective sensor data utilisation and analysis. In this review, we provide a taxonomy of more and less popular sensing modalities currently preferred by the experts in the field. These sensors are used to achieve target tracking and localisation. Additionally, we provide an overview of favoured fusion mechanisms employed to achieve higher accuracy [182], efficiency [83], robustness [57], or combination thereof [78]. The utilisation of these sensors is highly dependent on the use case. For example, there exist scenarios, where the accuracy of location estimation assumes a secondary role to energy efficiency [79] or user identification [53]. The selection of which sensors to use, and how, is usually left to user's preference and experience. This makes the relative selection space large, and frequently open to interpretation with regard to available resources and constraints.

 Whilst the survey literature pertaining to localisation systems and methods is large [54, 171, 181], there exists very little in the way of localisation-centric sensor utilisation. This encompasses the use of bespoke [39] or off-the-shelf [13, 118] sensors, specifically for the use of location estimation, robustification and optimisation. This area is extensive [20, 35, 57, 83, 136, 146], yet very often bundled along with localisation technology surveys, without subsequent scrutiny. We aim to close this gap, by reviewing sensors, their fusion and utilisation as applied to localisation, in contrast to localisation methods, technologies and implementations themselves.

Most of the existing localisation surveys include technology-specific reviews [30, 54, 92, 181]. They concentrate upon the methods and algorithms related to indoor localisation [30, 92], techniques and technologies [181]. Some work also addresses localisation from the perspective of the device itself, such as smartphones [171]. Xiao et al. study [171] is the most closely related work to our proposed examination. The main difference is, that instead of reviewing the devices as...
sensor clusters, we review the sensor modalities themselves. We also offer a more comprehensive review of fusion methods and provide exhaustive examples for each case.

The main contribution of this paper is the inventoryisation of the popular types of sensors used to provide location estimation and their respective advantages and disadvantages. We also provide the detailed description of their fusion methods with respect to their benefits and drawbacks. Finally, we show how these sensors are likely to fare in the future, paying close attention to the current community preference and trends surrounding each modality. To the knowledge of the authors, this is the first survey of its kind.

In Section 2 we outline the problem of localisation and provide a brief synopsis of the review process, concentrating on the most important indoor localisation-centric challenges found in literature. Then, in Section 3, we consider various sensors which are used in the service of localisation. In Section 4 we outline how the sensor fusion is performed, and review the state-of-the-art literature pertaining to effective sensor fusion and combination methods. Then, we provide a summary of the above inventoryisation in Section 5, and formally outline the likely future avenues for research. Finally, we conclude in Section 6.

2. Objective and Evaluation of Indoor Localisation

In this Section, we outline the evaluation criteria used to scrutinise the existing literature. We then list and discuss the sensors which are popular.

2.1. Semantic understanding of Indoor Positioning and Tracking

There exist various interpretations of positioning, navigation and tracking under the umbrella term indoor localisation. For example, Van Haute et al. [152] stipulates that tracking and positioning are not comparable. Whereas positioning implies establishing the location of an agent, either at real time or offline, tracking would involve performing localisation based on previous known location data, effectively storing the entire navigational history of an agent. This carries an additional risk of privacy intrusion, as the historical data would expose an agent’s habits and previous locations [152]. We intend to adopt a similar mindset, regarding the naming conventions of navigational methods in this review.

In addition to the above assertion, we consider it necessary to address a common misconception with regards to the semantic meaning of indoor localisation. A catch-all term, it grew to signify localisation inside, regardless of whether the environment is accessible by doors or not. In this paper, we understand indoor localisation to be an epitome of technologies and implementations for localisation in an enclosed environment. Examples of few such environments range from, but are not limited to, residential abodes [13], commercial shopping malls [160], industrial halls and factories [73], hospitals [64] and natural formations, such as underwater caves [99]. Here we consider sensor combinations stemming from the necessities imposed by these environments.

2.2. The Task of Probabilistic Localisation

Formally, the task of probabilistic localisation can be encapsulated by considering the minimisation of error between the location prediction and its corresponding ground-truth. If the true location in d dimensions is given by \( x(t) \in IR^d \), and its prediction \( \hat{x}(t) \in IR^d \), then:

\[
\min(\sqrt{\sum_{i=1}^{d} (x_i(t) - \hat{x}_i(t))^2})
\] (1)

that is, minimisation of the absolute Euclidean error between the prediction and label. Whilst there exist other metrics of evaluation [78], Euclidean error is by far the most popular [97], and is used extensively throughout this study.

Simply put, an agent traversing an enclosed environment is being localised if its location or navigational history is estimated with respect to their previous position, performed actions or current sensor reading. This estimation usually takes place in 2- or 3-dimensional space. The agent is assumed to be able to access the entirety of the surveyed environment. The model, or algorithm, performing the estimation also has access to the description of said environment as well as the features explaining the agent’s actions. In the domain of sensor-driven estimation, agent’s actions and locations are described through the use of sensors, which the agent either bears on itself or is subjected to, when travelling.

Simultaneous Localisation and Mapping (SLAM) is just one of the open problems in localisation literature, but it clearly and succinctly explains the challenge. In a perfect, noiseless world, the robot would be able to localise itself based on the Dead Reckoning (DR) alone. Then, by using the pictures, it would map out the environment, effectively solving the problem, by providing a map, and a vector of locations it visited. However, due to various conditions it is subjected to, noiseless localisation is so far unattainable. Its wheels will drift, adding noisy readings to the model. Camera pictures can be subjected to occlusion and lighting effects, making direct comparison difficult. The environment itself can also be dynamic, which adds to the complexity of the problem, as, in the case of this example, the photogrammetric features used by
the robot can be shifted, moved or otherwise removed from the corridor. For explicit explanation of the above problem, we invite the reader to [29].

The above mentioned Camera-SLAM example can be considered representative of the general problem of localisation. The noise associated with this method also explains the possible drawbacks of sensor driven indoor localisation approaches well. It should also be mentioned that the paragraph above explains a small subsection of a large field of study that is SLAM, and that Camera-SLAM was chosen due to its relatively intuitive explanation of the challenge. There also exist various other approaches to SLAM, some of which can be found later on in this text.

The motivation of using various sensor modalities, and their fusion, stems from the above mentioned issues. So far, there is no one definite way of performing localisation, as various sensors present different advantages and disadvantages. Whilst camera is known as a very accurate tool for feature extraction, it does so at the cost of high dimensionality and complexity of the data it collects. There exist modalities, which reduce the need for such high dimensionality, but in turn provide coarser location estimation. This implies that leveraging computational cost and estimation potential, across all modalities is, at the present moment, key to a successful implementation of an Indoor Positioning System (IPS) in GPS-denied settings.

### 2.3. Evaluation Criteria

The existing surveys of current localisation literature usually scrutinise the research through the use of a evaluation framework. Here we list the most popular criteria established either through literature [30, 54, 92, 181] or the authors’ own experience. This list is not exhaustive and is only provided to encapsulate the issues faced by the present-day implementations. Note also, that not all of these metrics can be applied to all of the scrutinised localisation methods and their utilised modalities. These will be used as evaluation as applicable.

**Distance Accuracy.** The most prevalent of metrics regarding localisation. Accuracy is usually calculated as Euclidean distance in 2D or 3D space [97]. Formal example is provided in Eq. 1. While effective, this metric is not infallible - there exist sensors and systems where a direct comparison of location accuracy (alternatively accuracy error) would not capture all necessary information required to examine any two given sensing systems. This point also considers whether certain sensors make it possible to scale the system to include more than one tracking node at a time.

**Noise resilience.** Sensing, in any form, will suffer from noise. This noise can be inherent in the sensing modality [88], environment [165], can be introduced during the manufacturing process[12, 112], or as a consequence of other factors, such as striving for improved energy efficiency [38]. Resilience of a sensor can also dictate whether drift and quantisation affect the location estimation and whether dependence on other sensor modalities can reduce it.

**Cost.** The costs associated with specific sensors are varied. These can be simple hardware costs, upkeep costs, deployment costs or maintenance costs. Hardware and upkeep costs encompass the initial expense of creating the infrastructure. Deployment and maintenance costs are related, in that they describe the value of labour associated with aforementioned tasks. Since different sensors will be comprised of different concessions regarding their performance and operation, they will all enjoy various advantages unique to their topology.

**Energy efficiency.** Efficiency has been cited as an important aspiration of a sensor-based system [37]. Deploying any system will come at a cost of establishing a number of trade-offs. Energy is often traded for accuracy/resilience to noise, as they tend to be mutually exclusive [125]. It is also important to recognise how easy is it to control the energy expenditure as part of a positioning system, and also whether the sensors make the system adaptable for energy-aware operation.

**Popularity.** The systems present within the literature rarely exhibit the same taxonomy of sensors, share the same evaluation environment or training methods. There exist implementations of positioning systems which consider various sensor modalities, and various fusion combinations. Currently, localisation relies on objective-specific sensor fusion as to ensure appropriate redundancy during its operation. The trends in literature are also greatly influenced by the relative costs and availability of hardware. We additionally aim to provide a future trend which the sensors are likely to take.

### 3. Sensor Modalities for Indoor Localisation

#### 3.1. Inertial sensors

Inertial sensors use the relative change in their frame of reference to provide an output. They are commonly employed in motion tracking and detection systems [43]. In relation to robotic or human localisation and tracking, they mostly comprise of Micro-Electro-Mechanical Systems (MEMS) accelerometers and gyroscopes, embedded within Inertial Measurement Unit (IMU) chipsets [167].

Accelerometers calculate the acceleration in 3-dimensional space, the domain of which is provided by black arrows in Fig. 1, given by units in g or...
They are particularly popular as part of Inertial Navigation Systems (INS), which constitute the focus of many localisation-centric research enquiries. There is a large body of literature pertaining to inertial sensing for localisation [2, 12, 67, 183].

As the manufacturing costs of MEMS devices reduced over years, their usage and the quality of their output has correspondingly increased. Lately, implementations feature smartphone devices which have these sensors readily embedded. One such study by Strozzi et al. [141] utilises a number of different hand held smartphones as a proxy to estimating step and its length. Similarly, Yin et al. [177] considers smartphone-based sensing, albeit as a tool for walking and running detection using accelerometers and gyroscopes embedded within.

While smartphones remain the favourite platform for sensing in many cases, there exist dedicated devices, so called wearables, which can provide acceleration and angular rotation from different parts of the body [10, 38]. Signatures from different sections of the human body were found to differ both, in the way they are exerted and their own estimation potential as per Bao et al. [10]. In our own study [78] we considered wrist-worn accelerometer as a complementary source of information in indoor location estimation. This method aimed to robustify the localisation performance by assuming that humans have a tendency of performing similar tasks in similar places in a house.

This type of sensing is not without its challenges however, as there has also been some advances in residential user identification. McConville et al. [104] showed that due to uniqueness of each person’s gait patterns, it is possible to recognise them directly from the inertial signals. The authors argued, that even though this was useful in pervasive health environments, it posed a significant privacy intrusion risk [104]. Off-body inertial sensor usage has also been investigated. Dang et al. [28] used different walking canes with attached IMUs to establish gait of the users, and consequently the distance travelled. This however relied on the participant using the cane with no abnormal deviations.

### 3.2. Ultrasonic and Acoustic Sensors

Ultrasound has also been explored for indoor localisation applications [56, 110, 119, 120, 184]. The basic implementation considers a number of speakers in the environment, which exert ultrasonic vibration [56] or frequency chirping [110]. The sensor designs themselves do not differ much from generic transducer-based microphones and speakers. In fact, this is done by using a piezo-ceramic or piezo-film transmitter, excited to generate a response at frequencies in [110] or over the human audible range [56], which is subsequently registered by a receiver.

The bulk of the localisation estimation is done through lateration schemes, such as Time-of-Arrival (ToA) [119, 121] and Time-difference-of-Arrival (TDoA) [110, 114] or angulation, like Angle-of-Arrival (AoA) [120]. They can be further categorised into Active and Passive [110]. Due to their physical nature, the sound waves experience similar shortcomings as electro-magnetic (EM) waves, in that they are limited by the Line-of-Sight (LoS) conditions However, when not experiencing multi path fading effects and
Figure 2. Example of tri-axial Accelerometer (2a) and RSS (2b) time series data from an indoor localisation ‘living’ experiment. Courtesy of Byrne et al. [13]
Non-Loss-of-Sight (NLoS) conditions, the localisation based on acoustic signal reportedly outperforms radio frequency (RF) based methods [110].

Early approaches, such as Cricket [119] used a combination of an ultrasound and RF to obtain a cheap localisation system. The experiments included static and mobile performance of the algorithm in an indoor office environment. This was later expanded into Cricket Compass [120] aimed at using angle of arrival in order to perform localisation.

More recently, Murakami et al. [110] used a smartphone-based mixture of active and passive signals. They were able to track the target along an open corridor. Qi et al. [121] used a number of ultrasonic receiver and transmitter modules in an Wireless Sensor Network environment. The aim was to establish a viable method for localisation under Non-Line-of-Sight conditions. This was tested by using a mobile robot, traversing in circles.

In their paper, Khyam et al. [75] used orthogonal ultrasonic chirping to utilise the wider part of the spectrum and facilitate multi-transmitter positioning in a passive context. Their experiments were carried out in largely noise-saturated environments. In the domain of robotics for indoor localisation Ogiso et al. [113] used a robot-mounted microphone array to attain positioning information of a pre-defined track. The robot would move in an 6m × 6m arena enclosed by four sources of sound, achieving sub-meter performance.

3.3. Visible Light Sensors

Visible Light Communication (VLC) is a subset of optical telecommunications concentrating on the visible light spectrum, or 380 to 780 nm wavelengths [127]. It supports faster transmission speeds [68], and offers a relief to congested radio frequency spectrum communication schemes [132]. Its fundamental operation relies on a source of light, such as a Light Emitting Diode (LED), modulated to flicker at a specific frequency, often to obfuscate the flickering. A light sensor is then used at the other end to receive and demodulate the transmission [132].

VLC is often used as part of the Visual Light Positioning (VLP) systems, whereby the modulated LEDs are used to estimate an object’s position, relative to lighting beacons [82, 131]. Much like Ultrasound, the schemes used to perform lateral or angular positioning rely on extraction of light signal strength [159] or relative AoA [82].

In their recent work, Rátosi et al. performed a real-time positioning based on LED anchor points [131]. In their work, they localised an object with a fish-eye lens camera extracting the positions and IDs of the LED beacons. They concluded that this approach is viable, even at relatively fast velocities of the object.

Wang et al. [159] was able to extract the beam strength of each uniquely-blinking LED through Fast Fourier Transform. Their LIPOS system was able to localise to within 2 meters Euclidean error in 3 dimensions.

Kuo et al. used a smartphone-based system to perform localisation, attempting to simulate the conditions usually found in retail spaces [82]. Their system considered using the lights mounted on the ceiling as beacons and smartphone’s front-facing camera as a capture method. Qiu et al. [123] used a kernel-based method to estimate the modulated light intensities. The authors noted, that due to the relative low-cost of the system and re-usability of an already existing lighting infrastructure, it could be used as a practical and efficient localisation implementation in the future.

3.4. Radio Frequency Sensors

This is undoubtedly the most examined area of indoor localisation implementations. RF-based sensing and location estimation have been the cutting edge methods of positioning due to their relatively low cost, off-the-shelf sensor availability and solid performance. This, coupled with the recent advances in Internet-of-Things (IoT) and ever-decreasing costs of maintenance have made this type of sensing a go-to for many researchers [8, 11, 13, 49, 78, 79, 103].

Whilst the number of technologies and standards within this group is vast, the basic idea of localisation remains the same. Generally, there exist a number of static anchor nodes, or Access Points (AP), which are able to transmit signals to a sensor traversing an environment of interest. They are comparable with ultrasound and visible light in the way that they are able to utilise similar schemes such as ToA and TDoA. Traditionally, Received Signal Strength (RSS) between a transmitter and a receiver was used as a metric to obtain information about the relative distance between the two nodes. This is made possible, as signal strength, assuming perfect propagation medium and lack of multi-path fading, will follow a steady decrease as a function of distance and is more formally described in terms of a path-loss equation [175]:

\[ PL(d) = PL_0 + 10n\log \frac{d}{d_0} + X_\sigma \]  

where \( d \) is the measure distance, \( PL_0 \) is a measured average path loss at a reference distance \( d_0 \) and \( X_\sigma \) is a zero-mean Gaussian random variable simulating the fading effect. This model is only an approximation of an indoor environment however, as the signal will vary in different surroundings and even different users [31].

A more realistic example is provided in Fig. 2b. There, the actual signal is obfuscated in noise, brought on by shadowing effects and fading. Recently, there have been
some work done using Channel State Information (CSI) [147, 175]. Using newer standards, such as IEEE 802.11, one can extract the amplitude and phase information from the channel directly, offering better performance [147].

The actual performance of RF localisation is deep-rooted in the technologies which are utilised to achieve it. Wi-Fi [42, 144] has been cited as one of the more popular approaches. Increasingly, the Bluetooth Low Energy (BLE) based sensors have been used, which leverage the low-power consumption with cheap cost and ubiquity [13, 150]. Radio Frequency Identification (RFID) [31] and Ultra Wide-band (UWB) [47] have also been used for location estimation, with UWB achieving sub-metre accuracy.

**Fingerprinting.** These schemes often rely on fingerprinting to achieve its performance. This consists of users visiting all fiducial locations in the environment, in order to build up an RF map [13, 178]. Whilst effective, fingerprinting has been recognised as difficult to obtain and maintain [13, 78, 79]. There have also been some work done, with multi-user environments, where it was confirmed that fingerprinting from one user is unlikely to be optimal on a different user [31]. There are however approaches designed to mitigate this difficulty [79].

The work done on RF localisation by Bahl and Padmanabhan [8] is widely regarded as the seminal paper on the subject of RF-based localisation. There, the authors outlined basic procedure for fingerprinting, where each required sector of the environment was characterised before outlining their algorithm for signal strength localisation. They used a specially fitted wireless adapters. Since then, the literature pertaining to sensor-based RF localisation steadily grew and so did the availability of off-the-shelf implementations.

Byrne et al. [13] presented a data collection of four different residential houses in Bristol. Each house was parametrised using approximately $1m \times 1m$ states, which permeated the living space. Then, a thorough fingerprinting of each abode took place. The dataset also included living experiments, and was performed using the SPHERE-in-the-box infrastructure [118]. This included Raspberry Pi-based access points and a bespoke SPHERE wearable sensor [39].

Wireless fingerprinting was also tackled by Yiu et al. [178]. They provide a comprehensive overview of fingerprinting methods, noting the online and offline phases of the radio map generation. Offline phase specifies the actual map generation, as in [13], and online phase is the location inference given current sensor output, which in their case was a Google Nexus tablet. They then outline different fingerprinting modalities, such as parametric (using path loss models) and parameter-free (based on Gaussian Processes).

**Lateration and Angulation Schemes.** There are also methods based on lateration and angulation of the signal from the prescribed sensor locations [26, 162]. These methods would assume that the signal propagation characteristics of some environment of interest can be directly calculated, and their solutions used to predict agent's movement directly. The difference between lateration and angulation is the method of calculation of the position. Whereas lateration estimates the position with respect to the direct distance from the sensor nodes (for example ToA), angulation does so, but with respect to the angle (for example AoA).

In [26], the authors offer a method for lateration, whereby the calculation of relative distances from provided sensors can be used to position a user. The study compared the methods based on least squares lateration and simple lateration schemes, using a smartphone, showing considerable improvement in positioning accuracy. In [69], the authors used a
trilateration scheme, based on Wi-Fi signals in order to localise an agent using a smartphone. In this paper, the authors made a distinction between LoS and NLoS conditions, achieving sub-2m accuracy. A paper by Paterna et al. [16] gives a thorough formalisation of lateration, and provides its own scheme, which the authors named ‘weighted trilateration’. Validation includes experiments based on frequency diversity, Kalman filtering and lateration, with reported best accuracy of sub-2m for a moving agent.

Park et al. [115] performed 3-dimensional localisation based on triangulation scheme from BLE nodes. The author performed the experiment, whereby 4 BLE beacons would be placed in the periphery of the central node. The results show that the authors’ method is at least as good as the current methods used to perform 3D localisation in the community.

3.5. Magnetometer Sensors

Ambient Magnetic Field (AMF) Localisation was inspired by the migration tendencies of certain animals [55]. Many species sense the Earth’s magnetic field and use it to navigate [55]. This method uses the extraction of a varying magnetic field inside buildings, in order to build a map of the environment, i.e. fingerprint. These distortions in magnetic field come from ferromagnetic fluctuations caused by the building’s metal construction and general topology [55, 173].

MEMS magnetometers [23, 143] are the most commonly used sensors in service of indoor localisation, due to their relatively low cost and high sensitivity [59]. They are generally used along with accelerometers and gyroscopes as part of PDR implementations [67, 70] where they act as directional sensors. However, they can also be used to estimate the ambient magnetic field in a given location inside a building [55]. They work by estimating the Lorenz force [59], measured as a function of current and magnetic field, given by [89]:

$$F_L = I B_X L_Z$$  \hspace{1cm} (3)

where $B_X$ is the magnetic field in T, $L_Z$ is the length of the loop or a wire in m, and $I$ is the current through the wire, in A. This force generates a displacement of a suspended control weight [89], which can be measured through piezo-resistive or capacitive means. The magnetic field induces current in the wire, which in turn forces the loop to move. The red piezo-resistors at the end of the loop in Fig. 4 are used to calculate the relative deflection and in turn, the causing magnetic field strength. Comprehensive outline is given in [59] and [89].

Haverinen and Kemppainen [55] stipulated that these anomalies in a magnetic field could be utilised for localisation. A subject wearing a magnetometer on their chest would walk along a corridor, measuring the field. Whilst they first proved its viability in a single dimension, this was later extended to 2 dimensions by Navarro and Benet [111]. However, the latter study was not directly comparable, as it was done using a wheeled robot as opposed to a human subject.

The popular approach of fingerprinting was appropriated to magnetic fields by Chung et al. [23]. In their work, the researchers used an offline map against which the observations were compared. The magnetometer was again worn on the chest, and proved comparable to other approaches, such as WLAN and RADAR. Similar fingerprinting was done by Subbu et al. [143], who published a smartphone-based localisation technique called LocateMe. The authors exploited the mobile phone’s magnetic sensor in order to gather fingerprinting maps of the environment and stipulated that this approach is also able to distinguish corridors with high precision.

3.6. Camera-based Sensors

When discussing camera-based localisation, it is important to distinguish between approaches where the localisation is a priority [164], and methods which render location information as a consequence of other inference, such as personalised silhouette detection [53, 150]. Whilst wide-scale indoor localisation with cameras is yet to be attempted, there are plenty of vision based tracking methods which consider smaller spaces [14, 153, 164].

There are many implementations of camera sensors on the market today. Digital cameras are most frequently based on CMOS technology [44] or obtained through charge-coupled devices (CCD) [128]. They are analogue devices, in the way they produce a lattice of pixels excited by the visible light to produce electrical signals, which are subsequently amplified and processed. Owing to its topology, this data is high in resolution and dimensionality [153]. This, in the context of indoor localisation, necessitates a streamlined and latency-free connection to a reference database to compare against a calibration set [153, 164].
or a thorough dimensionality reduction study [53] in order to become viable.

Early studies consider localisation through stereo vision. By using a stereo vision sensor, Bahadori et al. [7] presented a method of tracking multiple people in crowded environments, by modelling the background and the people themselves. This work outlined the basic principle of multi-person tracking in an indoor environment and noted issues with tracking identification.

Numerous approaches consider smartphone-based indoor localisation [153, 164]. Werner et al. [164] proposed MoVIPS, a visual positioning system. In their work, the authors used a smartphone to take pictures of the environment and compare them to a training set, with server-side feature extraction based on Speeded Up Robust Features (SURF). Similar approach was attempted by Van Opdenbosch et al. [153], albeit with a larger emphasis on efficient data analysis (by modifying a Vector of Locally Aggregated Descriptors (VLAD)), with comparisons between lossless and lossy compression.

As the depth-sensitive cameras became more cost effective, the research enquiry shifted to RGB-depth (RGB-D) sensors. Using RGB-D cameras for tracking has been established for some time [140]. In their work, Song et al. provided a large public dataset of RGB and RGB-D based videos for object tracking. RGB-D cameras are also widely used for Simultaneous Localisation and Mapping (SLAM) implementations [36, 142]. In these dataset papers, the consecutive depth-perceiving images are compared in order to evaluate location and at the same time produce a map.

In [109], Muñoz-Salinas et al. uses cameras in order to perform real time landmark-based visual SLAM. Here the authors used a fiducial markers, in order to estimate the location within the environment. In [33], the authors used 20 Kinect cameras in order to perform tracking of multiple targets transiting various trajectories. This was done in conjuction with Wi-Fi collected through user-carried smartphones. The authors reported sub-meter accuracy even in scenarios of 10 or more users walking simultaneously.

### 3.7. LiDARs

Light Detection and Ranging (LiDAR) devices are used as part of popular data association methods in order to obtain the position of the agent. They perform tracking by detecting the immediate vicinity of the agent and comparing it to previous readings [170]. LiDARs used in context of indoor localisation are most commonly found in robotics [60, 77]. There, the LiDARs are used most commonly utilised to perform SLAM [77].

![Figure 5. Example of a bird's eye view of a room outline (left) with 2-dimensional laser ranging device. The noisy LiDAR 'returns' are shown on the right.](image)

Whilst theoretically, any part of the light spectrum can be utilised to perform ranging, laser are most popular [61]. The working principle is rather simple and relies on ToA schemes - a beam of laser is sent out from the sensor and is reflected off the environment. Then, the time it takes to return is calculated from that beam, establishing likely distance between the LiDAR and the obstacle [25].

The data produced by a LiDAR can be either 2- or 3-dimensional [61]. This data is most commonly referred to as point clouds, due to discrete granularity of the environment it produces. These point clouds are later used as descriptors of the indoor environment and most commonly used to perform SLAM [60], usually as part of scan matching techniques [60, 163]. This data is however high dimensional and requires large reserves of computational power to optimise [77]. As shown in Fig. 5, point clouds are also susceptible to environment noise and jitter, which additionally creates scan matching issues.

Some early approaches to LiDAR localisation used robots in indoor positioning scenarios [21, 129]. Chmelář et al. used a laser range finder sensor in order to localise a robot in an indoor office environment. They used a compensation method in order to reduce the aggregated error. Rekleitis et al. was one of the first to propose a multi-agent localisation with LiDARs. Whilst the mapping was performed using a sonar, the robot agents were tracking each other using the LiDAR, in order to compensate for odometry errors.

Modern approaches enjoy better LiDARs and more computing power, allowing for faster processing and more resolute mapping [117, 163]. Peng et al. used

<table>
<thead>
<tr>
<th>Study</th>
<th>Method</th>
<th>Performance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Werner et al. [164]</td>
<td>Smartphone no markers</td>
<td>1.25m-4.4m</td>
</tr>
<tr>
<td>Muñoz-Salinas et al. [109]</td>
<td>RGB-D SLAM</td>
<td>0.16m-0.6m</td>
</tr>
<tr>
<td>Van Opdenbosch et al. [153]</td>
<td>Smartphone modified VLAD</td>
<td>&lt;1m</td>
</tr>
<tr>
<td>Domingo et al. [33]</td>
<td>RGB-D Tracking with Kinect and smartphone</td>
<td>&lt;0.5m for multi user scenario</td>
</tr>
</tbody>
</table>

Table 1. Table of recent camera-based systems with their method and performance.
a novel scan matching technique to achieve robot localisation in an indoor environment. Based on this work, Wang et al. [163], performed a similar study. Note that the robot used in both of the above papers was a ground-based device. Lee et al. [85] has used a LiDAR, along with a Virtual Reality (VR) headset, to obtain high resolution positioning using a drone. This experiment was in part inspired by disaster management and designed for first responders as an aid for finding survivors.

3.8. Other modalities

The above list is by no means exhaustive. In the literature, there exist various other implementations of IPS, which utilise less popular modalities. An example of one such implementation include Seo et al. [134], which used an ultrasonic anemometer to complement the IMU on a mobile robot. Anemometers measure relative velocity of air. In the above study, the robot was moving through static air, which ensured no erroneous readings.

Some research has also included piezoelectric infrared (PIR) sensors. Luo et al. [96] used a lattice-like sensor, in order to track an agent through the environment, at the same time performing activity recognition. The study motivated the use of PIR sensors, by noting that they are relatively infrastructure-free, and are easy and cheap to deploy. There also exist some data sets, where the PIR sensors are included, such as Twomey et al. [150].

There also exist studies using the piezo-electric effect in order to obtain the location and activity information of the users. The study of 'smart carpets' done by Chaccour et al. [19] does not cite indoor localisation as its main objective. However, this implementation could be used for very coarse location estimation as well. In the above study, the robot was moving through static air, which ensured no erroneous readings.

3.9. Drawbacks and Modality Evaluation

The presented modalities all differ in terms of the data that is being captured, and they way they obtain these readings. All of their topologies offer advantages and disadvantages in the domain of indoor localisation. It is possible, that the inherent form of data which a given sensor produces can provide a more or less confident estimate of the user’s position in the environment. The sensors in this review have been shown to produce viable localisation mechanisms. However, there exist sensors (such as accelerometers) which are more likely to be used in conjunction with other modalities (such as cameras) due to the performance they are able to obtain in positioning problems. It is important to distinguish the usability of each of the modalities before a more thorough discussion is provided.

Inertial sensors, whilst cheap and relatively energy efficient, often suffer from degrading noise [12, 112]. This noise is usually rectified by the researchers, though meticulous planning and closely controlled experiments [12, 67, 182]. Results 'from-the-wild' indicate that these sensors, are much more effective when used as part of a wider family of activity recognition tasks [13, 31, 32].

Ultrasound and acoustic sensors offer great precision but only at short ranges and in LoS laboratory conditions [113, 122]. Interestingly, most of the studies included in this survey have indicated that aside from these shortcomings, ultrasound is mostly preferred due to its low-cost and ability to reuse already existing sensor infrastructure, such as smartphones [110].

The biggest issue with RF sensing for localisation is the labour associated with training and the unpredictable nature of RF signals in the environment. The topology of this sensor make it great for tailored applications [78, 118], but often fail to generalise to other environments, and even users [31]. In addition, whilst fingerprinting is a powerful training technique, it is often cited as a drawback in any RF implementation [13, 87].

One of the major drawbacks of camera-based systems is the large computational complexity [153, 168]. Additionally, these sensors suffer performance degrading occlusion and lighting effects [14]. High dimensionality has also been cited as an important consideration [53]. These type of sensors are likely to be omitted in favour of other modalities in IPS settings. Magnetic field sensing has been proven to be effective, but only in confined spaces, taking advantage of ferromagnetic effects brought on by buildings [173], and under controlled conditions [52, 55]. This type of localisation also suffers from fingerprinting issues [23, 158]. Localisation based on an AMF could still be considered emerging, leaving plenty of opportunity for further work.

Visible light sensors provide a very accurate 3-dimensional positioning results at the cost of big infrastructures and controlled experimental testbeds [131]. Additionally, NLoS conditions are difficult to negotiate with this type of sensors [5, 88]. Modulation of the light beam is another issue - it requires frequencies large enough as to prevent visible flickering, which has been proved to be detrimental to the user experience [88].

LiDARs are a great intermediary between high dimensional data and reliable efficiency. However, the sizes and cost of these devices are still considerable when compared to the costs of inertial or even RF sensors. They are also prone to environment noise and,
Table 2. Table of sensor modalities, evaluated using the criteria from Section 2.3. This table summarises the criteria of various sensing modalities, additionally giving the justifying references for each.

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Accuracy</th>
<th>Noise Resilience</th>
<th>Cost</th>
<th>Current Consumption</th>
<th>Popularity</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Inertial</td>
<td>1m-10m</td>
<td>Low</td>
<td>Low</td>
<td>100µA - 3500µA</td>
<td>Outstanding</td>
<td>[28, 65] [39, 157, 166]</td>
</tr>
<tr>
<td>Ultrasound</td>
<td>0.01m-0.1m</td>
<td>Medium</td>
<td>Medium</td>
<td>Varies with application</td>
<td>Moderate/Emerging</td>
<td>[110, 122]</td>
</tr>
<tr>
<td>RF</td>
<td>1m-10m</td>
<td>Low</td>
<td>Low</td>
<td>Varies with protocol</td>
<td>Outstanding</td>
<td>[8, 47, 78] [166]</td>
</tr>
<tr>
<td>Camera</td>
<td>0.01m</td>
<td>Mid to high</td>
<td>Mid to high</td>
<td>Usage &amp; processing</td>
<td>Prominent</td>
<td>[7, 86]</td>
</tr>
<tr>
<td>Magnetometer</td>
<td>3m-5m</td>
<td>Mid to high</td>
<td>Low</td>
<td>~ 300µA</td>
<td>Emerging</td>
<td>[23, 143, 166] [55, 89]</td>
</tr>
<tr>
<td>Visible Light</td>
<td>0.2m-5m</td>
<td>Low (LoS)</td>
<td>Medium</td>
<td>~ 5100µA</td>
<td>Moderate</td>
<td>[63, 123, 159]</td>
</tr>
<tr>
<td>LiDAR</td>
<td>0.01m-0.1m</td>
<td>Medium</td>
<td>High</td>
<td>~ 300µA - 1A</td>
<td>Prominent</td>
<td>[60, 61, 138]</td>
</tr>
</tbody>
</table>

since scan matching relies on DR and will aggregate error over time, requires additional optimisation steps to become viable [60].

These modalities have been tabulated in Table 2, and scrutinised against the evaluation criteria provided earlier in this section.

4. Sensor Fusion

The above sensors are popular within indoor localisation literature. There exist numerous reasons for using these particular sensors on their own. However, by introducing an additional modality, one can obtain more information about the environment or its dynamics [78, 83]. By not relying solely on a single modality, an IPS can enjoy a number of advantages, ranging from resilience [15], accuracy improvement [20] or energy-awareness [79, 83].

Whilst, theoretically, fusion of any sensors is possible, not every combination is convenient. The most popular combination in the domain of inertial sensing, for example, is the consolidation of accelerometer and gyroscope with magnetometers, in order to produce robust PDR systems [80]. Nowadays, the relative energy output of these type of inertial sensors is negligible, which makes these sensors a popular choice in low-power applications [31].

RF-centric localisation has also been improved with fusion [15, 57, 146]. The combination of sensors in this context is usually performed for location improvement, as realistically, pure RF can only provide coarse location estimation. Mostly this involves either predicting or compensating the RF prediction with an inertial measurement [57, 78, 136]. Fusion of RF and magnetic field strength for performance improvement has also been explored [106].

In terms of robotic LiDAR SLAM applications, the fusion is also performed using the robot’s own IMU and magnetometer, in addition to the LiDAR [81]. VLC positioning has also been complemented by an IMU [185], as has ultrasound [48]. In each case this provides accuracy improvement to the system.

The relative fusion between different sensor modalities are visualised in Fig. 6. These sensor fusion combinations are by no means exhaustive. They were picked on the condition of being current examples of fusions between these types of modalities. Likewise in Fig. 6, the fusion was visualised only to help expose gaps in the literature pertaining to sensor fusion for indoor localisation. The intention of these is to give the reader a good starting point for their own investigations.

In the following sections we will review the studies which used fusion for a specific purpose.

4.1. Objective-specific Fusion Combinations

Fusion for Robustness. Fusion for robustness entails combining different sensor modalities in order to make the performance more resilient to outside adversity. Considering indoor localisation as our main motive, this adversity can come in the form of network-wide interruptions [78], dynamicity of the environment [98] or hostile agents [130].

By utilising Particle Filtering (PF), Canedo-Rodriguez et al. [15] was able to fuse a number of different modalities together for a robot-based indoor localisation system. These systems included LiDAR, Wi-Fi signal strength, cameras and magnetic signals from inside a museum. This robustification ensured a steady performance even in the event of dynamic environment, such as body shadowing. Li et al. [90] presented a technique for the fusing of UWB and IMU signals. This was done in the context of robotic indoor localisation.
using a Kalman Filtering (KF). The authors tested the algorithm against Gaussian noise, where their fusion method proved to be a viable safeguard.

Elbakly et al. [35] considered the fusion of a barometric sensor with Wi-Fi signal strength to provide a reliable prediction of floor transitions. It was tested thoroughly across three different environments, using 4 participants, and was proven to provide a robust performance across users. He et al. [57] used a Bayesian Network approach to fuse Wi-Fi and IMU signals. The authors arrived at the conclusion that the IMU was able to robustify the positioning based on a smartphone application.

Fan et al. [41] robustified the result of an DR-based indoor pedestrian localisation system using novel Kalman filtering and the fusion of MEMS-IMU. Through the use of robust fusion filter, the authors were able to reduce the overall aggregated error. This particular study additionally utilised a wavelet denoising method, as a preprocessing step, in order to remove as much inherent MEMS sensor noise as it was possible.

In the domain of robotics for indoor localisation, Paredes et al. [114] used a hybrid of an ultrasonic and camera-based sensing to achieve 3D positioning for an Unmanned Aerial Vehicle (UAV). The study concluded that purely ultrasonic localisation result is improved when using a ToA depth information from a camera.

**Fusion for Accuracy.** Accuracy in indoor localisation is most often calculated through the Euclidean error metric [97] and given in meters. Improvement of accuracy is the main ambition of many positioning studies. The fusion in this context would entail pin-point estimation of position based on a number of modalities. Over the years, many fusion attempts have achieved substantial reduction of positioning error, however no consensus among the community regarding the optimal way this fusion has to be attempted.

Similar approach to Canedo-Rodriguez et al. was attempted by Shi et al. [136]. The authors fused LiDAR and Wi-Fi, to robustify the accuracy of the location estimate. They compare a simple PF approach to their own, achieving considerable accuracy boost in a controlled environment. By using a KF, Chen et al. [20] fused Wi-Fi with landmark information on a smartphone sensor. In this study, the landmarks were found through unique locations of signature traces, such as elevators, stairs and steps. The authors were able to reduce the error of a single Wi-Fi based system by approximately 5m.

Yu et al. [180] performed the fusion of Wi-Fi and PDR on a smartphone, in order to achieve a better positioning accuracy of the model. They used an Unscented Kalman Filter (UKF) to provide a rough initial estimate of the location, before using accelerometers on the smartphone to estimate the location more precisely. The use of this system on an experimental track yielded considerable localisation accuracy improvement.

Zhang et al. [182] considered the fusion of a variety of sensors to achieve improvement on localisation using PDR, where the user was asked to take a challenging route up and down the stairs. Knauth also considered a PDR application [76] using the fusion of inertial, magnetic and RF sensors through a particle filter. It was again proven, that an inertial-based sensor fusion with Wi-Fi is able to outperform simple Wi-Fi-based positioning. Xing et al. [172] used the fusion of inertial, ultrasonic and optical flow sensors, along with ArUco markers in order to improve the positioning of a small drone.

**Fusion for Energy Efficiency.** In order to ensure continued operation of an IPS, the system itself has to be made aware of its energy usage. This is because the use cases of IPS usually necessitate them being operational for prolonged periods of time. Some of the implementations use smartphones as the computational foundation of their systems [76, 110]. Smartphones have been found to be less efficient than tailored implementations [83].

Kwak et al. [83] presented a system, based on the fusion of various inertial sensors and magnetic fingerprinting in order to achieve energy efficient IPS. The authors claimed a lifetime of almost a year on a single coin battery, at the same time reporting an error of 1.6m in a controlled office environment. Sung et al. [146] considered a smartphone-based inertial and RF fusion. In this work, the efficiency comes from the novel fusion implementations provided by the authors, and is validated with a thorough study of computational complexity between algorithms.

In our own work [79], we considered the utilisation of various sensor modalities for energy efficiency, using a Reinforcement Learning approach. Here, we were able to fuse BLE RSS with passive infrared and camera sensing to provide performance improvement over time, whilst retaining energy-awareness at all times.

### 4.2 Methods of Fusion

Having established possible reasons for fusion, we now consider the theoretical interpretations of the fusion methods which were previously mentioned. This subsection covers various generative and discriminative algorithms which make the fusion possible. They are listed in the order of their relative complexity.

**Bayesian Networks.** Bayesian Networks are often used in order to obtain a fusion of sensors [1, 139]. In a broad sense, Bayesian Networks are a subset of directed acyclic Graphical Models. The nodes of the graph
Sensor Modalities and Fusion for Robust Indoor Localisation

represent random variables which are being modelled. In a multi-sensor setting we can assume that the connections between the nodes in the graph represent their conditional dependencies. In other words, given a set of nodes \( x \), the general form of the joint probability distribution between variables is given by [139]:

\[
p(x_1, ..., x_N) = \prod_{i=1}^{N} p(x_i | Pa(x_i)) \tag{4}
\]

where \( Pa(x_i) \) are the parents of the node. Hidden Markov Models (HMM) are a popular example of dynamic Bayesian Networks, which are used to evaluate temporal processes. Found often in literature, their principle if rooted in the Markov property. They are formalised through the following equation:

\[
p(x_{1:t}, z_{1:t}) = p(x_0) \prod_{i=1}^{t} p(x_i | x_{i-1}) p(x_{i-1} | z_{i-1}) \tag{5}
\]

The equation above describes the overall process of evaluating joint probability between states \( x \) and observations \( z \) as a function of prior probability \( p(x_0) \), emissions (i.e. likelihood) \( p(z_i | x_i) \) and transition dynamics \( p(x_i | x_{i-1}) \). For further reading, we refer to [124].

There are many examples of Bayesian Fusion in sensor fusion literature [57, 62, 78]. He et al. [57] considered an HMM approach to fusion of multiple modalities on a mobile device using different graph structures for online and offline processing phases. Our own work, also based on HMM [78] involved scrutinising a number of different data flow models, which fused RSS and accelerometer data for robustness.

Hoang et al. [62] used a Bayesian approach to fuse RSS and steps detection signals for indoor localisation. The fusion proved superior to methods based solely on RSS. Similarly, Han et al. [50] used a novel approach to Viterbi coding to fuse RSS, Magnetic field and IMU traces to obtain an improvement on positioning accuracy.

Particle Filters. Particle Filters or Sequential Monte Carlo (SMC) are a subset of Bayesian Estimation methods. The basic algorithm relies on recursive estimation of the posterior probability of the state \( x_k \) given some sensor observation \( z_k \) at step \( k \). The objective of this algorithm is to estimate a probability density function associated with state \( x_k \), taking into account all sensor observations up to step \( k \), given by \( z_{1:k} \) [6]. This is done by first providing the prediction about our belief of \( p(x_k | z_{1:k-1}) \) and then updating the probability using Bayes’ Theorem. More formally [6]:

\[
p(x_k | z_{1:k-1}) = \int p(x_k | x_{k-1}) p(x_{k-1} | z_{1:k-1}) dx_{k-1} \tag{6}
\]

which is the prediction given by the Chapman-Kolmogorov equation [6]. The update can then be given by:

\[
p(x_k | z_{1:k}) = \frac{p(z_k | x_k) p(x_k | z_{1:k-1})}{p(z_k | z_{1:k-1})} \tag{7}
\]

Simply put, particle filters approximate probability density function of an unknown state as a recursive function of sensor observations which were observed up to some time. This particular approach has found applications in sensor fusion literature ranging from robotics [107], to activity recognition [133].

In the field of indoor localisation, they are most popular among the fusion of inertial sensors, especially when applied to PDR [3, 66, 126]. Hsu et al. [66] considered the fusion of a foot-mounted IMU and GPS signals to rectify noise drift. A similar approach was proposed by Akiyama et al. [3], albeit without the use of a GPS. There, the PF was scrutinised against energy efficiency, in addition to positioning accuracy. Racko et al. [126] also used particle filtering in service of PDR. They did this by predicting steps and heading from an IMU.

Kalman Filters. Kalman Filters are intimately related to recursive Bayesian filtering [34]. The popularity of KF was mostly thanks to its formulation, which allows many different sensor modalities to be arbitrarily modelled by the filter [46]. It is also preferred for its ability to obtain the result in real time. The usual KF formulation follows a pattern of state-space modelling, and their subsequent prediction and update [34].

Formally, the Kalman filter equation for state space input and output responses, in continuous time, are given by [34]:

\[
\dot{x}(t) = F(t)x(t) + B(t)u(t) + v(t) \tag{8}
\]

\[
z(t) = H(t)x(t) + \omega(t) \tag{9}
\]

where \( \dot{x} \) is the state vector, \( z \) is the output vector, \( u \) is the control input, \( v \) is the process noise and \( \omega \) is the noise due to measurement. Additionally, \( F \) specifies system state matrix, \( B \) is the input matrix and \( H \) is the matrix specifying the observations. The usual KF approach has two phases, prediction and update, which we will omit in our formalisation and instead refer the reader to [34, 46].

There exist work in the use of KF for indoor localisation [81, 134]. Kumar et al. used a KF to provide a 3D localisation of an indoor UAV, by integrating a LiDAR and an IMU. Here, the authors used KF to fuse the output of two LiDARs together to achieve 3-dimensional localisation.

KF can also be used as part of Extended Kalman Filtering (EKF), which is the nominal method used
in literature. EKF is a non-linear formulation of the KF, whereby the models of state transition can instead approximated through linearisation [148]. There exists a body of work dedicated to EKF for indoor localisation [18, 174]. Caruso et al., for example, uses an implementation of an EKF to perform localisation based on Visual-inertial Navigation System (VINS). They achieved superior performance to DR-based methods.

There is also a dedicated SLAM approach called EKF-SLAM [148]. In their paper, Vivet et al. [156] used a line-based EKF-SLAM for a robot based application. D’Alfonso et al. [27] also used an EKF-based approach to SLAM for a robotic indoor navigation task, supporting their simulated results with subsequent real life experimental work.

By using EKF, Alatise et al. [4] performed fusion of a 6 degrees of freedom (DOF) IMU sensor. They fused accelerometer and gyroscope to obtain the pose of the robot, i.e. the heading and its location. Kaltiokallio et al. [72] compared the relative performance of PF and EKF. The study concluded that for indoor positioning based on RSS, they are largely similar with the exception of the computational overhead, which favours the EKF.

Neural Networks. Due to the emergence of Artificial Neural Networks (ANN) in the recent years, a number of researchers have considered the use of a tailored network for sensor fusion. Most of the approaches use Deep Neural Networks (DNN) [84, 94, 161]. While there exists a body of literature dedicated to objective-specific fusion methods using ANN [151, 154, 155, 176], there is an evident lack of standardisation between the positioning methods, and it still remains largely unexplored.

Interestingly, ANN has often been used as a pre-processing step before actual fusion [154, 155, 161]. Whilst not strictly related to indoor positioning application, Vargas-Meléndez et al. [154, 155] used an ANN to estimate the pseudo roll angle of a vehicle, before performing fusion based on a PF. Wang et al. [161] performed indoor localisation, using CSI and deep learning. They were able to extract the location features by weighting them, using an ANN. This was later fused together during an online phase of their algorithm.

Liu et al. [94] proposed using deep learning for scene recognition and fingerprinting tasks. Using a smartphone, they were able to perform scene recognition from pictures using deep learning. Based entirely on the deep learning architecture, Lee et al. [84] performed localisation based on ambient magnetic field. They extracted magnetic features, as well as odometry and fed them to the network to obtain a robot’s position.

5. Future Directions

Figure 6 shows the fusion combinations and popular approaches in sensor-driven indoor localisation in the last decade. This particular figure is not exhaustive, and as it was noted before, is only attached as a starting point for further investigation of a particular fusion combination. Indeed, there is an evident community preference towards sensors which, either have a broad foundation on which to build the algorithms such as RF, or are based on modalities which are easy to come by, such as IMUs and magnetometers. While magnetometers have seen extensive use as part of PDR applications where they usually establish direction, there is lack of recent, comprehensive study of its viability with RF sensors. Both types utilise fingerprinting as part of its training phase. This type of data could be collected simultaneously, and can often reuse already existing IMU chipsets reported in various studies.

Cameras have seen a large body of literature dedicated to localisation, mainly due to the rise of camera-enabled smartphones. With easy access to smartphone sensor clusters, and their processing plants, researchers can perform more in-depth fusion of the sensors and collect more resolute data. Additionally, phones have good connectivity capabilities making them well suited for applications with quick-transfer requirements such as databases and for range-based RF localisation tasks. Interestingly, due to the recent trend in smartphone photography, where in order to obtain more resolute images the devices include two cameras, it could technically be possible to perform structure-from-motion mapping using a single smartphone with two or more camera sensors.

In terms of modality fusion, Ultrasound and VLC could both be considered relatively unexplored. Most of the literature, for both of these modalities, present implementations in a sterile environment of a laboratory, reporting sub-meter accuracy. That would suggest that these types of modalities are still in the proof of concept stage of research. There is yet to be study which would use these modalities in a wide-scale positioning infrastructure or fusion campaign. On the other hand, the fusion of RF and Inertia/Magnetometers is very widely explored, in both performance studies and their appearance in various data sets. The aforementioned Ultrasound and VLC-based approaches are, however, again underrepresented in this domain. This is not surprising due to the relatively large infrastructures demanded by these modalities. Additionally, there exists space for localisation-specific data set encompassing human-borne LiDAR for fingerprinting applications. This could be used with AMF or RF.
Fusion methodologies are also likely to shift. Recent proliferation of DL techniques and ANN in general, is likely to drive the fusion into the deep learning domain. Indeed, this paper has shown that there have been strides made in that direction, however when compared to Bayesian methods, this particular domain is lacking, in both proper theoretical formulation and exhausting comparison studies. This is not to say, that the current state-of-the-art Bayesian methods will be completely ousted. A more likely prediction is one of the two systems working together, either in unison, or as complements of each other, in order to make the prediction more accurate.

6. Conclusion

In this paper, we have reviewed the popular sensor modalities which are currently being used for indoor localisation. First, we have detailed each sensor modality and have given a thorough literature overview for each. The modalities were then scrutinised under widely accepted evaluation criteria. Then, we outlined the recent attempts at fusion and the most popular combination of sensors, considering context-specific consolidations. Among them were Robustness, Accuracy and Energy Efficiency. Finally, we have considered the popular sensor fusion methods, which range from Particle to Kalman Filters.

7. Acknowledgements

This study was supported by EPSRC and SPHERE IRC grant number EP/K031910/1. The authors would like to thank all colleagues from SPHERE and EurValve.

References


Sensor Modalities and Fusion for Robust Indoor Localisation

06. 004


