ZHEJIANG UNIVERSITY PRESS IET The Institution of Engineering and Technology WILEY

REVIEW

A review of smartphones-based indoor positioning: Challenges and applications

Khuong An Nguyen[1] (ID) | **Zhiyuan Luo**[2] | **Guang Li**[3] | **Chris Watkins**[2]

[1]School of Computing, Engineering and Mathematics, University of Brighton, Brighton, UK

[2]Computer Science Department, Royal Holloway, University of London, Egham, UK

[3]Institute of Cyber-Systems and Control, Zhejiang University, Hangzhou, China

Correspondence

Khuong An Nguyen, School of Computing, Engineering and Mathematics, University of Brighton, Brighton BN2 4AT, UK.
Email: K.A.Nguyen@brighton.ac.uk

Abstract

The continual proliferation of mobile devices has encouraged much effort in using the smartphones for indoor positioning. This article is dedicated to review the most recent and interesting smartphones-based indoor navigation systems, ranging from electromagnetic to inertia to visible light ones, with an emphasis on their unique challenges and potential real-world applications. A taxonomy of smartphone sensors will be introduced, which serves as the basis to categorise different positioning systems for reviewing. A set of criteria to be used for the evaluation purpose will be devised. For each sensor category, the most recent, interesting, and practical systems will be examined, with detailed discussion on the open research questions for the academics, and the practicality for the potential clients.

1 | INTRODUCTION

Indoor positioning is the technology that helps to locate objects or guide people in unfamiliar, complex buildings.

Despite its helpfulness, it has been almost 2 decades since Global Navigation Satellite System (GNSS) was introduced for outdoor positioning, yet the search for an equivalent ubiquitous, affordable, and accurate indoor counterpart is still going on. The challenge for such technology stems from the complex indoor interior design, and the building materials which block and distort the radio and satellite signals.

For the past 5 years, the proliferation of the mobile devices, along with the continuing miniaturisation of sensors, has propelled the smartphone as a potential instrument for future indoor positioning systems. Mobile phones have now eclipsed desktop computers in terms of worldwide market share.[1] For the developers, smartphones are not just mini-computers, but are also sensing devices with full awareness of their surroundings, providing a complete yet compact computing package. For the consumers, they need not buying new hardware, nor having to carry an extra device just for the positioning service.

Since 2005, the number of research papers involving smartphones indoor positioning has steadily increased (see

Figure 1), which implies the overwhelming interest in this topic. At the time of writing, there have been over 30,000 related papers indexed by Google Scholar since late 2007 when the first iPhone and Android were released.

Hence, to cater to the research community, there has been a wealth of indoor positioning surveys in the past decade, such as in [1,2]. However, many of them are either too broad by covering many general technologies including impractical ones, or too narrow by focussing on a certain niche market using proprietary devices that most consumers do not possess or cannot afford.

Therefore, this article is dedicated to the emerging smartphones-based systems, which inherit a massive user base as well as the high level of interest from academic researchers. At the end of the article, the authors aim to answer the following research questions.

- *What is the most promising, practical smartphones-based system for mass adoption?* This review will judge the system's practicality on the same set of criteria.
- *What are the most interesting open questions for each system category?* Academics working on smartphones-based indoor positioning may find interesting novel ideas for future research.

[1](https://gs.statcounter.com/platform-market-share/desktop-mobile/worldwide—last accessed in 5/2020)
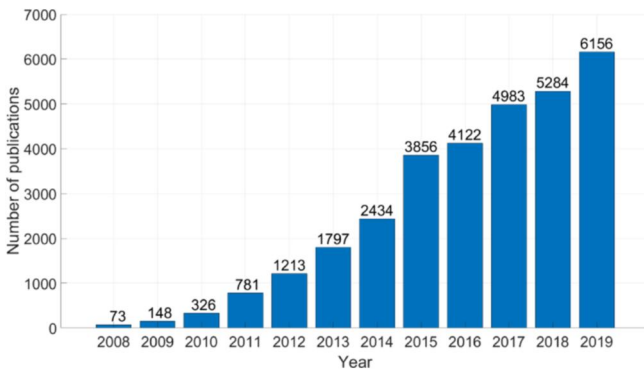
FIGURE 1 The estimation of the number of research work on smartphones indoor positioning indexed by Google Scholar. The search criteria were that the paper must contain both the 'indoor' and 'smartphones' keywords, and either 'navigation' or 'positioning' or 'localisation' or 'tracking', within the text. The results were then filtered by the authors based on their relevance

1.1 | Article's contributions

With the growing interest in the indoor positioning area and the high number of relevant research outputs, it would be beneficial for the research community to have a review article every 3 to 5 years. In addition, our article offers the following contributions.

- Only smartphones sensor-based systems were reviewed (i.e. ultrasound, laser scanner, etc. which are not present on the smartphones, are omitted). All sensors are discussed from the smartphones' perspective, and how they may be adopted for the positioning purpose.
- We emphasise on the practicality of each technology, and the open research questions.
- We devise a taxonomy to group different smartphone sensors.

1.2 | Style and structure of this review

To cater to the readers with interest in different systems, this article is written in a balanced narrative and systematic fashion, where a detailed, comprehensive literature survey is carried out, yet it is structured in a coherent, well-formatted template, allowing the readers to pick up any chapter independently. The mathematics is kept to a minimum, with only well-proven formulas describing the foundation of the technique included.

For each sensor category, we describe the general underlying technology, followed by how it is used specifically on the smartphones, including the provided measures and the challenges. At the end of each category chapter, we review the positioning methods that are applicable to this class of sensor, from a theoretical viewpoint with open research questions targeted at interested academic researchers. Then, we compare the performances of those techniques implemented on real-world systems as reported in the literature, aimed at practical implementations.

The remaining of this article is structured as follows. Section 2 introduces the concepts of smartphones-based positioning. Section 3 devises a taxonomy of smartphone sensor categories along with the criteria to review them. So that, Sections 4–9 will review individual category in detail. Last, Section 10 summaries the article and outlines further work.

2 | SMARTPHONES-BASED INDOOR POSITIONING

This section introduces the concept of smartphones indoor positioning, highlighting the applications and challenges facing such systems.

2.1 | Tracking with smartphones

For this type of system, we assume that the user must carry a smartphone with them at all times, in order to deliver the positioning service. This assumption is justified, given that 3.5 billion people possessing a smartphone in 2019, and over 94% of adults in the United Kingdom have one, according to a recent survey.[2]

However, smartphones do not come ready with indoor tracking capability. Thus, most solutions rely on an app running in background to deliver such service. This app receives data from the built-in sensors, some of which require the user's permission at launch time. Some systems may require additional supporting infrastructure to interact with the phone app, as we will discuss in detail in the upcoming sections.

Compared to other indoor positioning competitors, smartphones-based system offers unique benefits for the users in its compact form, and its self-contained package as a mini computing unit with an interactive touch screen. Most importantly, the clear advantage is being a ubiquitous device, the users need not carrying an extra piece of hardware for indoor positioning service.

2.2 | Challenges

Despite the aforementioned advantages, it is worth emphasising that most smartphone sensors, apart from GNSS, were originally designed for other functionalities in mind, rather than purposely for indoor positioning. Hence, we face the following challenges.

- *The sensor design is minimal*. Over the years, sensors continue to be miniaturised in order to fit into the small phone body. The challenge is that their sensitivity drops

---

[2](https://www.statista.com/statistics/300378/mobile-phone-usage-in-the-uk—last accessed in 5/2020)

consequently (i.e. a bigger sensor will have a longer range with more accurate measures).

- *The measures are noisy*. This includes the mechanical noises as the sensors are packed closely together in tight space with other conductive components like battery, and unintended interference from external sources.
- *The sensors are heterogeneous*. The variety of smartphones designers, and sensor manufacturers, makes it challenging to normalise the measures across devices.
- *The interior design is complex*. The building material is made of large ferrous metal structures (e.g. metal bars, steel rebars, reinforced concrete) which greatly distort some sensor readings.
- *The indoor environment is dynamic*. The building is usually occupied by many users moving about in their daily work, which may impact some wireless signal-based systems.

2.3 | Applications of smartphones indoor positioning system

Overall, most indoor positioning systems have two general purposes, tracking and navigation. For the tracking purpose, at any moment's notice, the system must be able to detect and pinpoint the location of the person or object. For the navigation purpose, the system must guide the users using the most optimal route. Table 1 compares some specific applications of indoor positioning for the four popular societal areas, which are healthcare, retailer, industry, and security.

3 | CLASSIFICATION OF SMARTPHONE SENSORS

Having discussed the general idea of smartphones-based positioning, we are now in a good position to delve deeper into the technical details of each individual sensor.

In short, a sensor is an electronic device that measures the changes in electrical or physical signals, and produces a measurable digital response to those changes. These changes can either be internal reactions within the mobile device, or externally in the surrounding environment. There are currently 14 sensors in modern smartphones (see Table 2 and Figure 3). Some of which are more useful for indoor positioning than others. In particular, we identify five properties that are most relevant for our purpose.

- *Temporal variation*. This is the sensor's ability to produce consistent measures over time, under the same environmental setting (e.g. in the same location).
- *Spatial difference*. For indoor positioning, it is critical that distinct locations have distinguishable sensor signatures, which are useful for pattern matching algorithms.
- *Battery consumption*. The lower the energy consumption, the longer the positioning app may continue running.
- *User permission*. Some sensors such as inertial ones can be queried at any moment, without a consent request, which enables a smoother user experience. Others require special

TABLE 1 The applications of smartphone indoor positioning system

Tracking purpose	Navigation purpose
Security	
For the guards:	*For the guards:*
• Being notified if the precious object has been moved.	• Finding the shortest way to the stolen/lost objects.
For the clients:	*For the clients:*
• Automatically granting access for authorised personnel.	• Quickly evacuating personnel during an emergency.
Healthcare	
For the patients:	*For the patients:*
• Locating the nearest wheel-chair.	• Route-finding to the doctor's office.
• Automatic checking-in/out upon entering the hospital.	• Automatic wheel-chair navigation in the hospital.
For the doctors:	*For the doctors:*
• Locating medical equipment.	• Finding the shortest route to the patient in an emergency.
• 24h monitoring of mental illness patients.	
Retailer	
For the customers:	*For the customers:*
• Shopping assistant for the latest in-store offers.	• Finding the correct shelf for a particular item.
For the managers:	*For the managers:*
• Sending out relevant e-coupons when the customers are near a particular shelf.	• Improving store layout by analysing footfall and congestion.
• Pinpointing staff's location in realtime.	
Industry	
For the customers:	*For the customers:*
• Tracking luggage at the airport.	• Finding seats in large venue.
	• Personalised computer guided tours.
For the administrators:	*For the administrators:*
• Locating products in the warehouse.	• Automatic adjusting the speed of conveyor belt for transporting heavy items.
• Be notified when an item leaves the warehouse.	

attention from the user in the form of a pop-up window to grant access (e.g., WiFi, Bluetooth, GNSS).

- *Sampling rate*. This metric determines how frequently the sensor updates its measure. A high sampling rate sensor is critical for tracking fast moving users and objects.

However, reviewing so many different sensors individually would make it inconvenient for the readers to follow, not least

TABLE 2 Comparison of the characteristics of 14 most common smartphones sensors, in alphabetical order, for the indoor positioning purpose

Sensor type	Temporal variation	Location accuracy	Battery consumption	User permission	Max. frequency	Description
Accelerometer	Low	Various	Low	None	196 Hz	Measuring the changing rate of the device acceleration
Ambient light	Low	Low	Low	None	4 Hz	Measuring the ambient visible light's intensity
Barometer	High	Low	Low	None	90 Hz	Measuring the atmospheric pressure
Bluetooth	High	Medium	Low	Yes	Various	Communicating with nearby Bluetooth beacons or other Bluetooth-enabled devices
Camera[a]	Various	High	High	Yes	60 Hz	Capturing the scenery via the front or back lenses
Cellular	Low	Low	High	Yes	20 Hz	Communicating with nearby cell towers
FM[b]	Low	Low	High	Yes	N/A	Receiving information from nearby radio towers
GNSS	Low	Various	High	Yes	10 Hz	Receiving the satellite signals to compute the latitude and longitude of the phone
Gyroscope	Low	Low	Low	None	198 Hz	Measuring the changing rate of the device's tilting angle
Magnetometer	Low	Various	Low	None	49 Hz	Measuring the ambient magnetic field strength
Microphone[a]	High	Medium	High	Yes	48 Hz	Capturing the ambient acoustic noise
NFC	Various	Various	High	Yes	1 Hz	Communicating with nearby RFID tags
Proximity	Low	Various	Low	None	4 Hz	Measuring distance to the nearest object within 10 cm
WiFi	High	Medium	High	Yes	0.03 Hz	Communicating with nearby access points or other WiFi-enabled devices

Note: The sensors are surveyed from the mid-range Lenovo Phab two phone.

Abbreviations: FM, frequency modulation; GNSS, global navigation satellite system; NFC, near-field communication.

[a]These sensors must operate in the foreground at all time.

[b]These sensors may not be available in all devices.
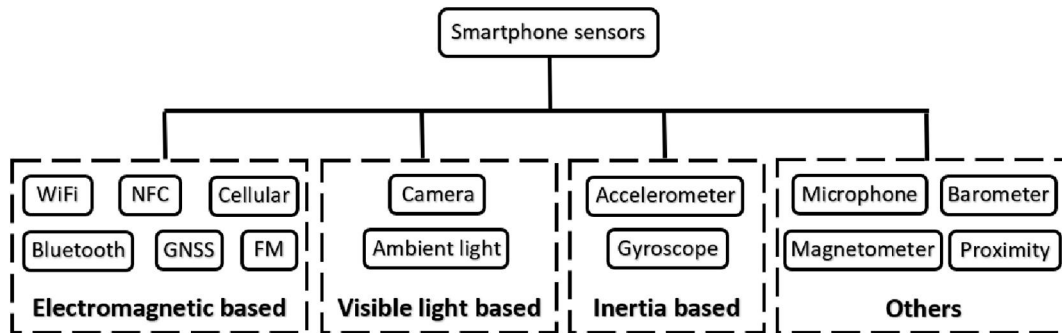
FIGURE 2 The taxonomy of smartphone sensors. FM, frequency modulation; GNSS, global navigation satellite system; NFC, near-field communication

some of them sharing common properties, and many techniques may be applied to multiple sensors. Hence, we derive a taxonomy to categorise those sensors into four groups based on their functioning mechanism, which are electromagnetic-based, visible light (VL)-based, inertia-based, and other sensors (see Figure 2).

- *Electromagnetic-based*. These sensors operate on the electromagnetic spectrum, through the invisible waves of wireless signals.
- *Visible light-based*. These sensors rely on visible natural lights to function.

- *Inertia-based*. These sensors use motions to estimate the phone's position.
- *Others*. This group includes the microphone, fingerprint, barometer, proximity, and magnetometer, which do not directly fit in the above groups.

The sensors from each group will be reviewed together in their own separate section. In each section, we will overview the technologies behind each system category, from the smartphones' perspective, emphasising on their challenges, the positioning algorithms, and the results of the most notable systems in the recent literature.
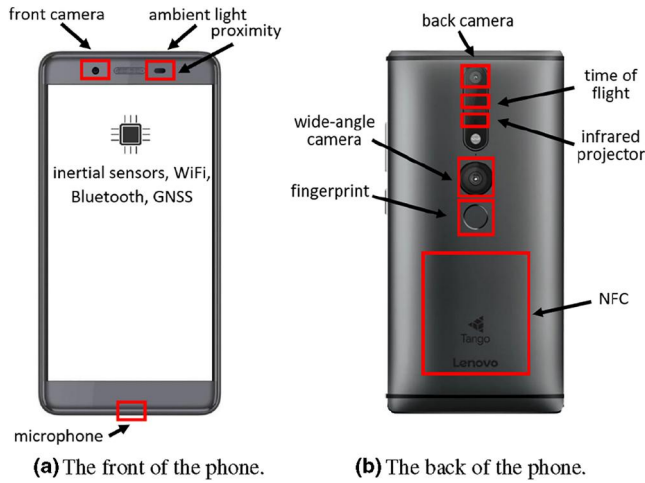
FIGURE 3 Relative position of most smartphones sensors, illustrated on our Lenovo Phab 2 test phone. Some sensors are embedded within the phone, while others are exposed. The battery occupies the majority of space at the bottom half of the phone. (**a**) The front of the phone. (**b**) The back of the phone

4 | ELECTROMAGNETIC SIGNAL-BASED SYSTEMS

This section examines the systems that operate on the electromagnetic spectrum, with frequencies ranging from as low as 10 Hz to the crowded 2.4 GHz band (see Figure 4). These electromagnetic waves carry information, generated by both electric and magnetic components, across the space without the need of any transport medium, and can be interpreted by the smartphone sensors. Some systems are ubiquitous as the transmitters already exist for other purposes, while others require dedicated hardware deployment, to be discussed later on.

Distance wise, the lower the frequency of the system is, the longer the wavelength is (i.e. the further the signal can travel, and the easier it can pass through walls, furniture). This attribute demonstrates why low frequencies are reserved for (AM), frequency modulation (FM), Cellular transmitters, which need to reach long distance users. In contrast, the higher the frequency of the system is, the smaller the wavelength is (i.e. the shorter distance they travel). To compensate for the lack of distance, these systems have a higher bandwidth capacity to serve many users simultaneously, which explains why WiFi, Bluetooth transmitters are on this frequency.

Data wise, electromagnetic signals are capable of carrying position-related information from the transmitters (e.g. the MAC address, the wireless channel, geo-coded data, etc.) that the receiving smartphone may infer to determine its location.

4.1 | WiFi

WiFi describes the technology that allows the local devices to communicate wirelessly via an access point (AP) on the 2.4 GHz band, or the newer 5 GHz band. It was designed to replace the wired Ethernet cable for faster and more secured data transmission. Modern buildings are populated with plenty of APs for Internet access and wireless communication amongst the users, which inherently benefits WiFi-based systems as no extra hardware is needed.

On the smartphones, WiFi was intended to allow the user to connect to nearby hotspots or routers for Internet access, or the recent WiFi Direct protocol that allows two smartphones to exchange data directly without an AP. Therefore, the level of information exposure is not as expansive as on a PC or laptop. In particular, from the smartphones' perspective, the visible information is just the name, MAC address, and the received signal strength (RSS) of nearby APs. To obtain this information, the phone initiates a scan, which listens on each of the 14 WiFi channel on the 2.4 GHz band for a short period of time for any data periodically sent by nearby APs. This process is known as passive scanning. By default, WiFi APs broadcast data simultaneously on all channels every 100 ms. Thus, the entire scanning process takes about 1.5 s on most smartphones. Other devices such as laptop or PC may initiate a faster active scanning process that floods the probe request frames on all WiFi channels, and waits for the probe responses from the APs. The process normally completes within 200 to 500 ms, which is significantly faster than passive scanning. Unfortunately, active scanning is not supported on Android devices to preserve battery and reduce overloading the WiFi channels.

Therefore, the challenges for WiFi RSS-based system are:

- *Indoor areas with similar RSS*. Since the APs were primarily designed with data communication in mind, their deployed positions may not be optimal for indoor positioning.
- *Changing indoor interior*. Any rearrangement of furniture may alter the way the WiFi signals are deflective, absorbed, and scattered around the building.
- *The missing APs*. As these systems rely heavily on the WiFi APs, an AP taken off-line may impact the positioning result, or worse, requiring a new survey of the building.
- *The scanning delay*. As discussed above, passive scanning on Android devices may take up to 1.5 s, which may not be suitable for monitoring fast moving users.
- *WiFi cell breathing*. This mechanism is implemented in most new WiFi network to permit overloaded APs with too many connected users to offload their work to neighbourhood APs. The consequence is that the coverage zone, and APs RSS, which are critical for a positioning system, dynamically change depending on the number of active users.

In 2016, the new IEEE 802.11mc standard introduced a protocol called fine timing measurement (FTM) which is also known as WiFi round-time-trip (RTT) [3]. This new protocol allows WiFi-enabled devices to measure the distance to nearby WiFi APs, by timing the elapsed time the signal travels from the device to the AP and back. The most significant benefit of
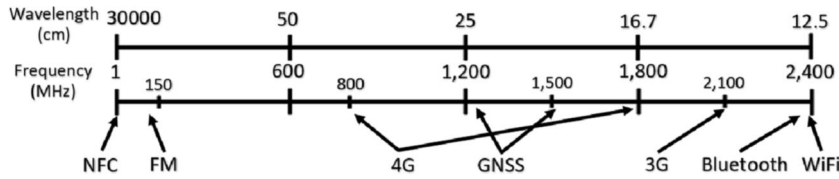
RTT is that it requires no clock synchronisation between the transmitter and the receiver. Hence, WiFi RTT was intended to be more robust and more accurate than RSS. However, WiFi RTT still faces the following challenges.

- *Limited support*. At the time of writing, only four WiFi APs (i.e. Google WiFi, Google Nest Router, Google Nest Point, and Compulab WILD AP), and eight Android phones, were reported to be working with RTT.[3]
- *Signal multipath*. Similar to RSS, RTT may be impacted by the signal multipath effect. The most accurate estimation may only be achieved when there is a direct line-of-sight between the two devices.
- *Various hardware and software implementation*. The difference between the WiFi chipset and the firmware may result in various processing time delay, which impacts the RTT measure.

It is worth noting that there are other types of WiFi-based distance measure, most notably is the Channel State Information (CSI), which addresses the major WiFi multipath challenge by providing rich, fine-grained information of the channel response, obtained from the PHY layer (whereas traditional RSS is obtained from the MAC layer which does not separate any multipath characteristics). However, the CSI measure is not exposed by the official Android OS, and requires proprietary hardware (e.g. Intel 5300 NIC) and custom software [4].

Overall, despite its challenges, WiFi is still perhaps one of the most popular choices for indoor positioning right now, due to its ubiquity.

4.2 | Bluetooth

Bluetooth describes the wireless technology that allows devices to communicate directly to each other, on the same 2.4 GHz band as WiFi. Although both Bluetooth and WiFi appear to facilitate the same wireless communication purpose among devices, their design intention is different. First, Bluetooth's target audience are small peripherals (e.g. mouse, keyboard) with a few concurrent connections (typically between only two devices), while WiFi's target is delivering high speed connection among many larger devices (e.g. laptop, PC). Second, Bluetooth devices are easy to switch on/off and be ready to

connect to others, whereas WiFi APs require excessive configuration. Third, Bluetooth's coverage is much shorter at 5–10 m, comparing to typically 30–50 m from WiFi's. Last, Bluetooth was designed for power efficiency and affordability (i.e. a single 1.5 V cell battery may power a Bluetooth beacon for up to 9 months), while WiFi's aim was to maximise high speed data transmission.

On the smartphones, Bluetooth was primarily aimed to connect the phone to low energy consumption peripherals such as headphone, fitness tracker, and to exchange small pieces of information where speed is a non-issue with other Bluetooth-enabled devices.

Hence, similar to WiFi, the level of exposure is rather limited, in the form of just the MAC address and the RSS of nearby Bluetooth devices, which are obtainable from a scan. However, the major difference to WiFi is the duration of the Bluetooth scan, which depends on three parameters—the scan window (i.e. the length of a single uninterrupted scan), the scan interval (i.e. the gap between scans) of the phone, and the advertising frequency of the peripheral device (i.e. how often the device announces its existence on the three Bluetooth advertising channels). Once setup, WiFi sticks on one channel throughout its life, whereas Bluetooth frequently hops between its three channels. Ideally, it is preferred to have a long scanning window with a short interval, and a high advertising frequency to maximise the discoverability of nearby Bluetooth devices. However, the consequence is the high power consumption for both the phone and the peripheral device. By default, the most optimal Bluetooth scanning profile on Android (i.e. the SCAN_MODE_LOW_LATENCY) sets both the window and the interval to 4096 ms. This means the phone will scan uninterruptedly for just above 4 s, with another 4 s break till the next cycle. Therefore, the Bluetooth peripherals should be configured to advertise at least once every 4 s to be discovered by the phone.

Since 2017, the latest Bluetooth 5.0 iteration introduced enhanced location services with the new direction finding feature (i.e. Angle of Arrival [AoA] and Angle of Departure [AoD]). With the AoA approach, the transmitter which often comes with just a single antenna (e.g. a tag, a beacon) transmits a special signal containing directional information. The receiver (e.g. a smartphone) is equipped with multiple Bluetooth antennas arranged in an array fashion, to capture such signal from different perspectives, which then estimates the relative direction of the signal. With the AoD approach, the transmitter-receiver roles are reversed, so that, the smartphone with multiple antennas transmits multiple directional signals, which are then captured by the signal antenna tag or beacon,

---

[3](https://developer.android.com/guide/topics/connectivity/wifi-rtt—last accessed in 9/2020)

to work out the relative direction [5]. It is worth reminding that the new Bluetooth 5.0 chipset is needed to facilitate such a function.

Nevertheless, the general challenges for Bluetooth based systems are:

- *Deployment of Bluetooth beacons*. Having to manually install and find the optimal placements for all beacons as well as maintaining these devices are challenging.
- *Signal instability*. Due to the implemented frequency hopping technique, the Bluetooth signals from a beacon often fluctuate even when observed from the same location.
- *High latency*. The 4 s interval between scans may pose a challenge for applications requiring constant location updates.

Overall, with its relatively short range and low power consumption, Bluetooth is suitable for beacon-based solution.

4.3 | Global Navigation Satellite System

GNSS describes the constellation of satellites circling the Earth in a precise orbit two times per day [6]. The GNSS sensor on the smartphones acts as a receiver to measure the distance from itself to the satellites, based on the receiving signal, and applies trilateration to determine the location of the phone (to be discussed in the positioning method section) using the satellites as reference points. In contrast to the above WiFi and Bluetooth technology, which were designed for the indoor environments, GNSS system was designed primarily for outdoor use.

From the smartphones' perspective, the available information are the phone's current longitude, latitude, timestamp, and accuracy (i.e. the radius of the circle with the phone at the centre), which are precisely what most positioning systems need. However, GNSS-based indoor positioning faces the following challenges.

- *Extremely weak satellite signals*. Theoretically, the satellite orbits are arranged in such a way that at least four satellites are seen at any place on Earth. In reality, skyscrappers may block such view. Additionally, by the time the satellite signals reach the Earth's surface, they become too weak to penetrate most modern building materials.
- *Coarse positioning*. In normal working condition, GNSS has an average of 5–10 m accuracy, which is sufficient for outdoor navigation. However, such accuracy means the user may be estimated to be in a different room or building.
- *Excessive battery consumption*. GNSS receiver on smartphones is known to drain battery quickly when used for a long period of time.

Overall, in the indoor context, GNSS is usually employed as opportunistic signal information, rather than a stand-alone indoor positioning solution.

4.4 | Cellular

Cellular technology allows mobile devices to call, text each other, as well as accessing the Internet via the mobile network provider. At the heart of this technology is the cell tower, which serves a specific geographical area (i.e. cell). When the users move between cells, their information are automatically handed over among towers. To avoid signal interference, adjacent cells use different signal frequencies.

From the smartphones' perspective, the available information are the cell ID, the signal strength, and the signal-to-noise ratio of nearby cell towers. Most notably, there are multiple large crowd-sourced cell tower databases,[4,5] which contain the geographical location of most towers, to be used for trilateration of the mobile devices.

Nevertheless, cellular-based indoor positioning faces the following challenges.

- *Highly coarse positioning*. It is common for cell towers to be placed 2 to 3 km apart, making it challenging to distinguish the user location within a cell.
- *Weak signal*. Although cellular signals operate on lower frequencies than GNSS, allowing it to penetrate most building materials to which the satellite signals struggle, well-shielded buildings with thick walls, glasses may still block out the cellular signals.

Overall, due to its low spatial signal difference, cellular technology may only offer city/town level positioning accuracy.

4.5 | Frequency modulation

Frequency modulation (FM), also widely known as FM radio, is the technology to broadcast sound-based information such as music, news to listeners with a receiver on the same frequency band. Each radio station occupies a fixed frequency on the FM spectrum, and has its own radio tower. Compared to cellular cell towers, FM radio towers are significantly more powerful transmitters, covering a much wider geographical area (see Figure 5).

The clear advantage of FM over other electromagnetic signals such as GNSS, WiFi, Bluetooth, Cellular, is that it operates on a significantly lower frequency, from 87.5 to 108 MHz, making it easier to penetrate buildings and walls.

Nevertheless, since FM utilises the same concept of outdoor broadcasting tower as cellular, it suffers from the same weaknesses, which the readers may refer to Section 4.4. The biggest challenge for this type of system is that Android currently does not officially support the technology, and most APIs are vendor-specific, making it challenging for mass market deployment.

---

[4](https://www.opencellid.org—last accessed in 5/2020)
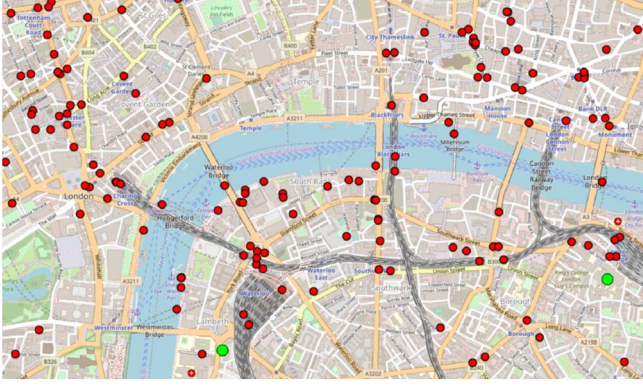[5](https://radiocells.org—last accessed in 5/2020)

FIGURE 5 The deployment of cellular towers of the UK O2 network and FM transmitters in central London. There are 141 cellular towers, depicted by the red dots, whereas there are only 2 FM transmitters, depicted by the green dots in the same area

4.6 | Near-field communication

Near-field communication (NFC) describes a set of close range wireless protocols, enabling two devices within 10 cm to communicate. The technology is based on inductive coupling on the 13.56 MHz frequency, between the antennas on NFC-enabled devices.

From the smartphones' perspective, NFC is made popular by contactless payment applications (e.g. Apple Pay, Google Pay), which allow the consumers to make payment on the go by tapping their phones onto a payment device. The technology also allows smartphones to read information (e.g. room number, geo-coded position) embedded on NFC tags attached on walls, doors, furniture. The advantages of this approach are that NFC tag is passive with no internal power, and is rather inexpensive to deploy.

However, NFC-based systems face the following challenges.

- *Short communication range*. The active working distance is just within 10 cm. Hence, the users must almost touch their phone to an NFC tag or device, which impacts the hands-free feature of the system.
- *Limited simultaneous reading*. Android allows only one tag to be read at a time, although the technology is capable of observing up to four tags stacked on top of each other, simultaneously.

Overall, because of the short working distance within 10 cm, NFC-based solutions may not be suitable for the indoor positioning purpose yet.

4.7 | Positioning algorithms

Having understood the strengths and weaknesses of each electromagnetic-based sensor, we are now in a good position to introduce the most popular positioning methods for these sensors, namely proximity tracking, trilateration, and location fingerprinting. At the end of the section, we will overview the most notable recent systems in the literature.

4.7.1 | Proximity tracking

This method aims at detecting whether the phone is close to a tracking beacon (e.g. the WiFi APs, Bluetooth beacons). It is suitable for systems in which only a binary decision about the presence of the tracking person in a certain area is needed (e.g. whether the customer is near a particular shelf to send out coupons, whether the user is inside a room to switch on the light), and not the precise fine-grained position. In these cases, the location of the person is assumed to be the known location of the tracking beacon (see Figure 6). The sensors that benefit the most from this method is short range ones (e.g. Bluetooth), as they only broadcast signals within a small area.

The advantage of such an approach is in its simple implementation, as there is no computation involved in. The disadvantage is clearly its limited positioning accuracy as the user is coarsely estimated within a wide area.

4.7.2 | Trilateration

Trilateration improves on proximity tracking's accuracy by combining the RSS from multiple nearby beacons. It does so under the assumption that the position of at least three nearby tracking beacons is known. This is the technique that underlines modern GNSS localisation [7].

It works in three steps. First, the phone measures the RSS to the nearest three beacons. Second, for each beacon, it draws a circle representing the whereabouts of the phone with respect to the beacon (i.e. a weak signal will result in a small circle). Last, the intersection of these three circles is the estimation of the phone's position (see Figure 7). Without loss of generality, assuming the 2D co-ordinate of the three beacons are (x_1, y_1), (x_2, y_2), (x_3, y_3), the distance d_1, d_2, d_3 between each beacon and the user is calculated based on the Pythagorean theorem as follows [8]:

$$d_1 = \sqrt{(x_1 - x)^2 + (y_1 - y)^2}$$
$$d_2 = \sqrt{(x_2 - x)^2 + (y_2 - y)^2} \qquad (1)$$
$$d_2 = \sqrt{(x_3 - x)^2 + (y_3 - y)^2}$$

The intersected user position (x, y) is then calculated as:

$$x = \frac{A(y_3 - y_2) + B(y_1 - y_3) + C(y_2 - y_1)}{2(x_1(y_3 - y_2) + x_2(y_1 - y_3) + x_3(y_2 - y_1))}$$
$$y = \frac{A(x_3 - x_2) + B(x_1 - x_3) + C(x_2 - x_1)}{2(y_1(x_3 - x_2) + y_2(x_1 - x_3) + y_3(x_2 - x_1))} \qquad (2)$$

where $A = (x_1^2 + y_1^2 - d_1^2)$, $B = (x_2^2 + y_2^2 - d_2^2)$, $C = (x_3^2 + y_3^2 - d_3^2)$.
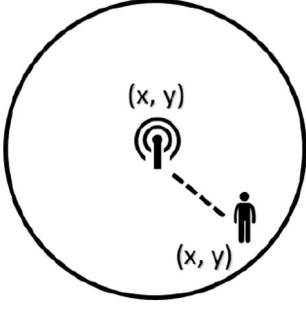
FIGURE 6 Proximity tracking. The user location is assumed to be the position of the nearest (strongest) beacon
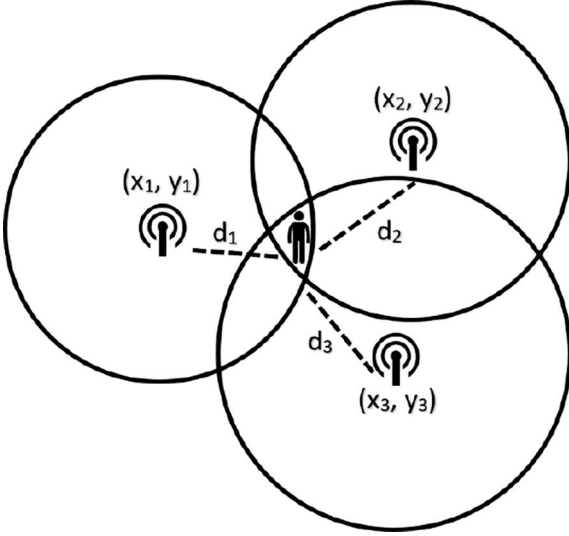

FIGURE 7 Trilateration method. The user location is inferred as the intersection of the nearest beacons using geometry

The distance between the user and each beacon d_1, d_2, d_3 can be obtained from the wireless signal strength (e.g. WiFi RSS) using several distance model, for example, the logarithmic distance path-loss model as follows [9,10]:

$$d_{1,2,3} = 10^{\left(\frac{TxPower-RSS}{10*n}\right)} \tag{3}$$

where TxPower is the signal strength measured at 1 m from a known beacon (e.g. −65 dBm), n is the propagation constant (e.g. $n = 2$ for free space path-loss constant) [11,12], RSS is the measured signal strength from the user position, and $d_{1,2,3}$ is the distance between the beacon and the user (in metre).

The pros of trilateration is its practicality with simple geometry calculations. Its biggest cons, however, is the indoor signal attenuation. While such technique works great outdoors for GNSS, with clear view to the satellites, there is often no line-of-sight from the phone to the indoor beacons. The signal is greatly attenuated that distorts the geometric shape of the broadcasting beacons which in turn does not provide accurate positioning estimation. A variation of trilateration is known as triangulation, which looks at the receiving signal's AoA to compute the position.

4.7.3 | Location fingerprinting

Fingerprinting is arguably the most researched method in the indoor positioning literature. It addresses the weaknesses of proximity and trilateration methods by eliminating the need for both line-of-sight and the location of the beacons. The core concept of fingerprinting is based on a training database, generated to capture the wireless signal at every location in the building. Thus, the more unpredictable and challenging the signal propagation is, the more unique the training samples will be, which helps estimating the position of a user at an unknown location using his wireless signal [13].

Fingerprinting has two phases, off-line surveying and online positioning estimation (see Figure 8). In the off-line phase, an expert surveys the building to generate a training database. He does so by walking through the building with a smartphone in hand to measure the wireless signal and the physical co-ordinate (x, y, z) at every position. In the on-line phase, when the user wishes to discover his position, he measures the wireless signal at his current position. The system then searches through the training database generated previously to find a best match. This process is also known as pattern matching [14,15].

Fingerprinting is generally considered a supervised machine learning (ML) problem, as the training database contains both the object (i.e. the wireless signal strength) and the label (i.e. the Cartesian position). For this type of learning, there are two main approaches to estimate the positioning label. The first approach uses the training database to create a model, either probabilistic (e.g. Naïve Bayes) [16,17] or deterministic (e.g. Ridge Regression, SVM) [18,19], which is then later used to estimate the position of a new signal sample [20]. In contrast, the second approach skips the model learning phase to go directly from the training samples to estimating the position (e.g. K-nearest neighbours) (see Figure 9).

Without loss of generality, the simplest form of fingerprinting can be formalised as follows. Given the training database with M training examples $T_i = (\overrightarrow{RSS_i}, \overrightarrow{L_i})$ $(1 \leq i \leq M)$, where $\overrightarrow{RSS_i}$ is the wireless signal strength, and $\overrightarrow{L_i} = (d_x^i, d_y^i, d_z^i)$ is the Cartesian positioning label of each training position i, the task is estimating the position $\overrightarrow{L_u} = (d_x^u, d_y^u, d_z^u)$ of a new wireless signal vector $\overrightarrow{RSS_u}$ somewhere in the building.

With the non-model-based approach, the training example T_i whose signal strength is nearest to the new sample's, is chosen as the estimated position. A popular measure metric for the RSS vector is the Euclidean distance:

$$Eu(\overrightarrow{RSS_i}, \overrightarrow{RSS_u}) = Eu((AP_1^i, ..., AP_N^i), (AP_1^u, ..., AP_N^u))$$
$$= \sqrt{\sum_{p=1}^{N} (AP_p^i - AP_p^u)^2}$$

$$(4)$$

where AP_i is the RSS measure of WiFi AP i, and N is the number of APs.
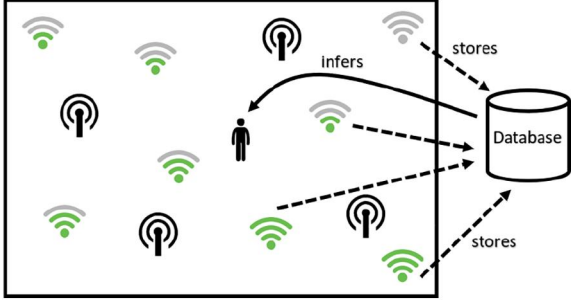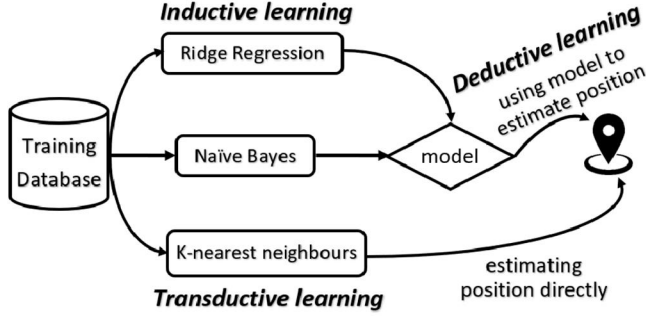
FIGURE 8 Fingerprinting method

FIGURE 9 The machine learning approaches for fingerprinting

However, as the WiFi signal attenuates, different positions may possess a similar RSS. Hence, the nearest training sample may not be the correct estimation. Thus, a set of K nearest training samples are often considered. This approach is known as weighted K-NN, where a weight parameter is employed to penalise anomalous training samples as follows [21–23]:

$$L_u = \frac{\sum_{i=1}^{K} \frac{1}{Dist(\overrightarrow{RSS_i}, \overrightarrow{RSS_u}) + \epsilon} \overrightarrow{L_i}}{\sum_{i=1}^{K} \frac{1}{Dist(\overrightarrow{RSS_i}, \overrightarrow{RSS_u}) + \epsilon}} \qquad (5)$$

where ϵ is a small positive constant to avoid division by zero.

With the model-based approach, we train an ML model using the provided training database. A popular algorithm is Ridge Regression which looks for a model that fits all training examples $T_i = (\overrightarrow{RSS_i}, \overrightarrow{L_i})$ $(1 \leq i \leq M)$ [24]. This model will later be used to estimate the positioning label $\overrightarrow{L_u}$ for the new RSS vector $\overrightarrow{RSS_u}$. Ridge regression is based on mean squared error (MSE), with an extra tuning parameter called the ridge factor $\lambda \geq 0$ used to control the balance between fitting and avoiding the penalty. The ridge regression model is presented as follows:

$$\mathcal{M} = \sum_{i=1}^{M} (\overrightarrow{L_i} - w \cdot \overrightarrow{RSS_i})^2 + \sum_{j=1}^{N} \lambda \|w_j\|^2 \qquad (6)$$

where w_j $(1 \leq j \leq N)$ are the weights for the N WiFi APs. The closer λ is to zero, the more linear the regression model is. In contrast, the bigger λ is, the more weights close to zero or equal zero are. The objective is to find the optimal weights w^* so that $\mathcal{M} \to \min$. The estimated position $\overrightarrow{L_u} = (d_x^u, d_y^u)$ for the new sample $\overrightarrow{RSS_u}$ is computed as follows, where each Cartesian dimension has a separated model:

$$d_x^u = w_x \cdot \overrightarrow{RSS_u}$$
$$d_y^u = w_y \cdot \overrightarrow{RSS_u}$$
$$d_z^u = w_z \cdot \overrightarrow{RSS_u} \qquad (7)$$

Another popular take to the model-based approach is modelling the training samples using a probabilistic model, in which the WiFi RSS are captured repeatedly at each training position to record the full signal distribution of each AP [25,26]. This model addresses the challenge of the WiFi RSS variation in the same position, causing a mismatch between the training RSS and the real-time RSS, despite being collected in the same place. Without loss of generality, given the new RSS sample $\overrightarrow{RSS_u}$, the algorithm calculates the posterior probability $P(T_i|\overrightarrow{RSS_u})$ of this RSS sample being observed at the training position T_i $(1 \leq i \leq M)$. The training sample with $\max_{i}(P(T_i|\overrightarrow{RSS_u}))$ is the estimated position. This probability is computed using Bayes' theorem [27]:

$$P(T_i|\overrightarrow{RSS_u}) = \frac{P(\overrightarrow{RSS_u}|T_i) \, P(T_i)}{P(\overrightarrow{RSS_u})} \qquad (8)$$

Since $P(T_i)$ and $P(\overrightarrow{RSS_u})$ are given, all left is calculating the reverse probability $P(\overrightarrow{RSS_u}|T_i)$. By assuming the WiFi RSS from each AP are independent measures, we may apply the Naïve Bayes approach to calculate the posterior probability for each RSS as follows:

$$P(\overrightarrow{RSS_u}|T_i) = \prod_{j=1}^{N} P(AP_j^u|T_i) \qquad (9)$$

where $P(AP_j^u|T_i)$ presents how often the individual RSS of AP_j appears at the training location T_i

$$P(AP_j^u|T_i)$$
$$= \frac{how \; many \; times \; AP_j \; appears \; at \; T_i}{total \; number \; of \; readings \; observed \; at \; T_i} \qquad (10)$$

The major pros of fingerprinting is it exploits the complex indoor environment as its advantage. The more unpredictable the signal attenuation is, the more unique the training data will be, which aids the positioning estimation process. However, the major drawback of fingerprinting is the sheer amount of building space to be meticulously surveyed, the time it takes to

perform such process, and the lack of physical reference for the training positions [28]. As fingerprinting relies on a training database, below are some open research questions.

- *How to alleviate the surveying process?* This is the most challenging procedure of fingerprinting, especially for a large building.
- *How fine-grained the training database should be?* The building is normally segregated into evenly spaced grid where the calibration points are collected. The denser the points are, the bigger the training database will be, and the more efforts need to be invested.
- *How to obtain the position training label?* Each signal data needs a positioning label (e.g. the Earth's latitude and longitude, the Cartesian co-ordinate), yet it is not trivial to obtain such data. Most systems rely on an external system for ground-truth positions. Others assign them manually.
- *How to cope with indoor interior changes?* As the fingerprinting database is a representation of the signal propagation in the building at the time of measure, any interior changes (e.g. furniture rearrangement) may require a complete re-surveying of the building.

4.8 | Performance review

Having discussed the most popular techniques in the last section, we are now in a good position to review their performances on real world systems. It is worth emphasising that this section only reviews single sensor-based systems to expose the true performance-related measure of each sensor (see Table 3). Hybrid systems will be reviewed later on in Section 9.

Technique wise, the overall consensus is that fingerprinting consistently achieved higher positioning accuracy at an average of 2–3 m [17,23,29], compared to others. Proximity- and tri-lateration-based systems, while being simple to deploy, could barely manage room-level accuracy at 5–6 m accuracy [29,30]. Nevertheless, despite its high accuracy at the time of testing, fingerprinting-based system may struggle to sustain such performance over a long period of time, because any environmental changes may require the entire training database to be re-calibrated, as discussed earlier. Additionally, the density of the training points as well as the number of beacons also contribute to the positioning accuracy [31–35].

Sensor wise, Bluetooth offers the highest accuracy, at 1.3 and 1.8 m, 90% of the time [36,37]. It is worth noting that these works strategically populated their own Bluetooth beacons densely for the training database. This observation was echoed by another work with the FM signal, where WiFi fingerprinting with more beacons, achieved double the positioning accuracy of FM fingerprinting, under the same testbed [38].

Overall, when it comes to system performance, WiFi and Bluetooth seem to be the two most popular options among electromagnetic sensors. WiFi, with its ubiquitous coverage in most modern buildings, offers slightly above room-level accuracy, whereas Bluetooth with its compact size, low power consumption, may offer 1–2 m accuracy, when manual deployment is possible.

4.9 | Overall remarks and open research questions

Electromagnetic-based approaches are perhaps the most researched category for indoor positioning. Its strength lies in

TABLE 3 Comparison of notable electromagnetic signal based system performance

Authors	Sensor	Positioning technique	Testbed	Accuracy	Notes
Zhao et al. [17]	WiFi	Probabilistic fingerprinting	60 × 20m	3.2 m, 90% chance	The authors slightly improved the positioning accuracy with filtering and AP selection.
Shin et al. [23]	WiFi	W-KNN fingerprinting	48 × 22m	5.3 m, 90% chance	Almost double the positioning accuracy compared to standard K-NN was reported.
Mok and Retscher [29]	WiFi	Trilateration	20 × 20m	4 m, 95% chance	Only five WiFi APs were deployed in a relatively small area.
Faragher and Harle [37]	Bluetooth	Fingerprinting	600 m^2	1.8 m, 90% chance	Dense deployment of one beacon every 30 m^2.
Nguyen [36]	Bluetooth	Fingerprinting	5 × 2.5 m	1.3 m, 90% chance	Dense deployment of five beacons in a small area. A robot was used to automate the surveying process.
Canton et al. [30]	Bluetooth	Trilateration	16.5 × 17.6 m	8.2 m, 90% chance	By exploiting the BLE channel diversity which is not available on the smartphones, the accuracy may be improved to 7 m, 90% chance.
Moghtadaiee et al. [38]	FM	W-KNN fingerprinting	11 × 23m	5.2 m, 90% chance	WiFi fingerprinting under the same testbed was reported with double the positioning accuracy.
Viel et al. [39]	GSM	Fingerprinting	Udine city centre	45 m, 90% chance	Different techniques including training models (Random Forest, SVM), KNN achieved similar positioning accuracy.

the flexibility in catering to both infrastructure-based and infrastructure-free systems. As most modern buildings are already populated with many WiFi APs, the developers can utilise them, along with other opportunistic signals such as Celullar, FM, GNSS, and the Earth's magnetic field [40]. For developers preferring to deploy their own tracking beacons, Bluetooth APs offer an affordable solution with low power, and minimal beacon design.

Positioning accuracy wise, systems in this category normally achieve around room-level accuracy on average. Maintenance wise, depending on the chosen technique, there may be overhead in updating the training database, servicing the tracking beacons.

Research wise, some interesting open questions for the academic researchers are:

- *Electromagnetic interference*. The 2.4 GHz band in which WiFi and Bluetooth operate, is overcrowded with devices (e.g. phones, laptops, PCs) and home appliances (e.g. microwave, cordless phone, baby monitor). Collision in such environment is inevitable, which results in loss of connection, or beacons not visible during a scan.
- *Signal attenuation*. There is rarely a clear line-of-sight between the phone and the beacons. Hence, the wireless signals propagate unpredictably in the air. When two in-phase waves of signal meet, constructive interference results in a stronger signal, whereas, a combination of two out-phase waves will result in destructive interference which weakens both [28]. The end receiving signal at the phone is a product of these phenomena. Besides, the user body which contains much water, is also a big attenuation factor [36,41].
- *Device heterogeneity*. Various antenna from different smartphones have different sensitivities which impact the receiving signal. This has a major impact on systems relying on RSS.

5 | VISIBLE LIGHT-BASED SYSTEMS

This section discusses the systems that rely on natural, VL to the human eyes, in contrast to the electromagnetic signal systems discussed in the last section, which use invisible waves of signal.

Distance wise, all systems in this category need a clear unobstructed sight from the sensor to the tracking person or object, as VL cannot penetrate walls and furniture. Perhaps, this is the biggest challenge for this type of system.

Data wise, VL does not physically carry information the way electromagnetic wave does. Light-based systems rely on passively analysing the unique features of the scenery (i.e. whether there are enough distinct optical information among places), or actively illuminating the scene with their own light source and capturing the reflective light back (i.e. to measure the distance to the nearest object). Some systems used custom LEDs that flicker rapidly in unique pattern to distinguish among themselves, which can be decoded by the smartphones camera, yet are indiscernible to the human eyes. By altering the flickering frequency, information (e.g. the co-ordinate of the beacon) may be encoded within. This technique is known as VL Communication (VLC).

5.1 | Camera

Smartphones camera is CMOS imaging sensor, which is designed for low power device to capture an image of the surrounding. It does so by converting light waves passing through the lenses into digital signals.

On the smartphones, camera was designed for either capturing scenery using the back facing camera, or taking selfie using the front facing camera. Hence, from the smartphones' perspective, the output is in the form of an image or video (i.e. a sequence of images). Modern smartphones are capable of recording up to 60 frames per second. The only assumption for camera-based system is that there must be sufficient ambient light for the sensor to generate a clear image reflecting the surroundings.

Nevertheless, camera-based systems face the following challenges.

- *Phone placements*. Since the camera is fixed at the back of the phone, different angles will produce various perspectives of the same scenery.
- *Lighting condition*. The image sensor easily becomes under- or over-exposed, under different lighting.
- *No distance information*. Although the camera is capable of observing the world around, the captured information contain no depth information (i.e. it has no idea how far away an object is).

Overall, smartphones camera offers a means to see the world the way human do. It is currently one of the most popular choices for optical-based systems.

5.2 | Time-of-flight

Time-of-flight (ToF) is a 3D image sensor used to measure the distance to nearby objects. It tackles one of the biggest challenges facing standard CMOS cameras described earlier in the sense that everything captured in the image appears as in a flat world without any depth perception. ToF works by illuminating the scene with laser pulses, and observing the reflected light (i.e. the faster the light bounces back, the shorter the distance is) [42,43].

On the smartphones, ToF has two main functionalities, to scan the user's face for phone unlocking (using the front facing sensor), and to blur the background for portrait photography (using the back facing sensor). The former is much more popular in modern smartphones. Technically, the returned information is a point cloud containing the z-dimension depth information. These 3D points can be meshed together to create a surface rendering of the object and the environment (see Figure 10).
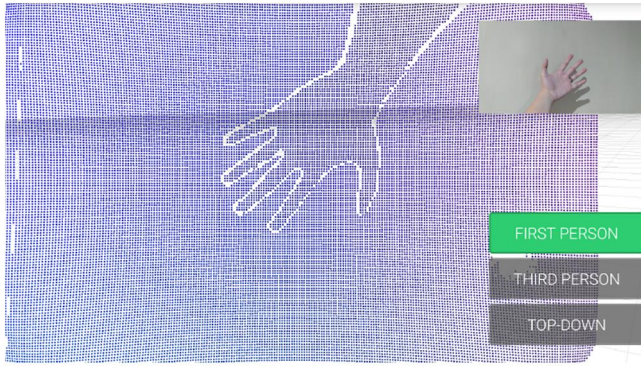
(a) Under normal lighting.
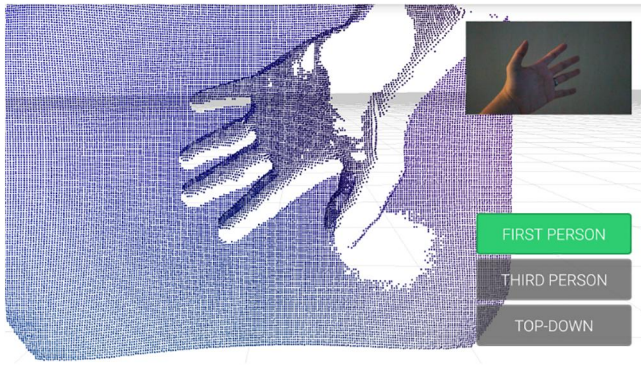
(b) Under dimmed lighting.

FIGURE 10 The ToF point cloud image of a hand in different lighting conditions, observed by our Lenovo Phab 2 phone. (**a**) Under normal lighting; (**b**) Under dimmed lighting

The big advantage of ToF is its high immunity to the ambient light, as it uses its own light source. However, the challenge for this type of technology is that only a handful of smartphones come equipped with this sensor. Additionally, ToF faces the following challenges.

- *Sensor placement*. Most smartphones only employ the front facing ToF camera for facial recognition purpose. This set up is unlikely to be used by indoor navigation app to read the depth information of the scenery facing the other way, while the user is looking at the screen.
- *Reflective and absorbing surfaces*. As the distance measure is based on the reflected LED light, shiny surfaces (e.g. metal, glass) or absorbing surfaces (e.g. wood, plastic) make the measure inaccurate.
- *Slow frequency*. While standard CMOS camera can capture images at 60 frames per second or higher, current ToF cameras are restricted to 5 Hz, which may pose a problem for high speed scenes.
- *Limited range*. Most ToF sensors cannot see beyond 4 m, as the LED beam becomes saturated beyond this range.

Overall, a combination of ToF and standard camera would closely resemble the way human perceives the world around. As more smartphones come equipped with this sensor, it is a promising technology for future indoor positioning systems.

5.3 | Ambient light

The ambient light sensor is a photodetector sensor, designed to detect the amount of light in the surroundings. It does so by converting the light energy into electric currents. On the smartphones, the sensor is exclusively used to adjust the screen's brightness with respect to the current ambient light.

The challenge for this sensor is the information it measures change significantly by various factor (e.g. the shadow of the human body, the facing angle of the sensor).

5.4 | Positioning methods

Having discussed the strengths and weaknesses of all VL-based sensors, we will now overview the most popular positioning methods, and the most notable recent systems in the literature.

5.4.1 | Proximity and trilateration

These methods are identical to the ones discussed for the electromagnetic-based systems. Hence, we refer the readers to Sections 4.7.1 and 4.7.2. In short, with proximity-based tracking, when a person enters a sensor coverage area (e.g. a CCTV enabled zone), her presence will be recorded. With trilateration, as each sensor's coverage range is known, and the final position is deducted as the intersection of these areas.

5.5 | Triangulation

The triangulation technique determines the location of the user by computing the angles in relation to at least two known fixed position beacons. In our case, the beacons are custom LEDs with unique flickering patterns, which are captured and decoded by the smartphones' front facing camera to work out its current position (see Figure 11). The working principle is inspired by the way mariners navigate on the sea, by observing the brightness from a set of fixed position stars in a constellation.

Compared to trilateration, which simply uses a rough estimation of the distance (i.e. WiFi RSS), triangulation is more knowledge-demanding as it requires not just the beacons position but also their spatial rotation. Additionally, because of the limited field-of-view of the smartphones camera, a dense LED grid needs to be deployed to guarantee visible beacons at any holding angle and position of the phone.

5.5.1 | Fingerprinting

The concept of fingerprinting for VL-based systems is identical to that for electromagnetic-based systems (see Section 4.7.3). Instead of using the wireless RSS, optical information are used to create the training database. As the smartphones only provide image captures as outputs for each

position, computer vision techniques were employed to extract notable features from such images. Ideally, these features must uniquely represent the position (i.e. no two positions should have the same set of features), as well as being robust to different lighting conditions. The most widely used technique is 'Scale Invariant Feature Transform' (SIFT) [44,45].

The crux of this technique is finding locations (i.e. key points) within an image space that are invariant to image scaling or rotation, as well as being least affected by optical noises and distortions.

Without loss of generality, given the image $I(x, y)$, the first step is defining a continuous function of scale known as scale space $L(x, y, \sigma)$ that is produced from the convolution of a variable-scale Gaussian $G(x, y, \sigma)$ as follows, with $*$ is the convolution operation and σ is the variance of the Gaussian filter [46].

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y)$$
$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/2\sigma^2} \tag{11}$$

To detect key points in this scale space, the difference of two nearby scales is computed as follows [44]:

$$D(x,y,\sigma) = (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \tag{12}$$

where k is a constant factor to separate the convoluted images in scale space.

For each image, each point is compared to its 17 neighbours in the current image and the ones above and below in the scale space. A point will be selected as a key point, if it is either larger or smaller than all of its neighbours.

5.5.2 | Visual odometry

This method attempts to estimate the position and orientation of a moving phone by analysing the sequence of images taken by the on-board smartphones camera. It does so under the assumptions that there must be significant overlaps between continuous images (i.e. two images taken several seconds apart on a fast moving user will not contain sufficient overlaps), and the scenery must contain distinguishable textures (i.e. a big empty room with plain walls does not have unique information from any viewing angle).

Visual odometry has four steps [47,48]. In the first step, the image is captured by the smartphone's camera, whose notable features are then extracted. In the second step, the features between two consecutive images are compared (i.e. using features matching) based on some similarity metrics. Another approach for this step is called feature tracking, that is only extracting features from the first image, then track them in the next image with some correlation techniques. In the third step, the camera motion is estimated by either a Kalman filter, or finding the geometric properties of the features that minimise the cost function of the two consecutive images. In the last
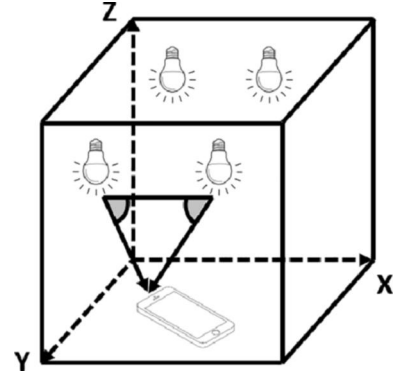

FIGURE 11 The triangulation approach with VL Communication (VLC)

step, a time window is applied to refine the walking trajectory of the last N frames.

Without loss of generality, the problem can be formulated as follows. Given a time-series of N images taken by the smartphone camera $I = (I_1, ..., I_N)$, the two camera positions T_k and T_{k-1} taken at two consecutive time k and $k - 1$ are related by the rigid body transformation as follows [48].

$$T_{k,k-1} = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix} \tag{13}$$

where $R_{k,k-1}$ is the rotation matrix, $t_{k,k-1}$ is the translation vector.

Finally, the camera poses are denoted as $C = (C_1, ..., C_N)$, where $C_i, 1 \leq i \leq N$ is the transformation of the camera at time i with respect to the initial position at 0. The current camera pose C_N is calculated by merging all transformation $T_k, 1 \leq k \leq N$.

5.6 | Performance review

Having introduced the most popular techniques for VL-based systems, we are now in a good position to review their performances on real world systems. The general impression is that most VL-based systems achieved significantly higher positioning accuracy at centimetre level, than the previous electromagnetic-based systems at metre level (see Table 4). More details will be discussed as follows.

Sensor wise, systems that use custom LEDs managed to achieve very high positioning accuracy at 22 cm and 40 cm, 90% chance, respectively [49,50]. However, it is worth noting that these systems employ dedicated LEDs that flicker at unique patterns. Realistically, it is unlikely that most venues will replace their existing lighting infrastructure with these custom LEDs just to support the indoor positioning service. It is preferable to deploy new beacons which are less disruptive, or ideally making use of the existing infrastructure. On this note, ToF system managed to achieve similar positioning accuracy, at 15.2 cm mean error [51], without having to deploy additional LEDs. Nevertheless, very few current

TABLE 4 Comparison of notable visible light based system performance

Authors	Sensor	Positioning technique	Testbed	Accuracy	Notes
Yang et al. [49]	Eight custom LEDs and camera	Triangulation	2.4 m × 1.8 m × 3m	22 cm, 90% chance	The authors employ polarisation-based modulation to enable a low pulse rate positioning system.
Li et al. [50]	Five custom LEDs and camera	Trilateration	5 m × 8 m	40 cm, 90% chance	Channel hopping was employed where each LED beacon transmits via the optical channel which was claimed to be ambient light interference free.
Zhang and Zhang [52]	Unmodified fluorescent lights and camera	Fingerprinting	9000 m², 119 lights	25 cm, 90% chance	The authors use normal lights in the building. Each FL has an inherent characteristic frequency which can serve as a discriminate feature. They claimed 10 cm, 90% chance while the user stands still.
Xiao et al. [53]	Camera	Fingerprinting	1000 m²	1.5 m mean error	The authors built a training database containing image of static landmarks i.e. doors, windows. Experiments were done in a large open space museum.
Fang and Scherer [51]	Time-of-flight	Visual odometry	64.2 m × 21.2 m × 3.9 m	15.2 cm mean error	The authors also tested the system in challenging dim environment with similar accuracy.

smartphones support ToF sensors, although this trend may change in the near future.

Technique wise, triangulation and trilateration, despite being simple to implement, excelled in this sensor category with very high centimetre positioning accuracy. However, these systems relied on additional hardware, as discussed previously, which adds the installation cost, and the disruption of setting up such hardware into the tracking area. On the other hand, fingerprinting-based VL systems, which maintains a training database of either the lighting characteristics of the environment [52], or the image of static landmarks [53], managed to achieve just above metre positioning accuracy. These systems conveniently fit into any building without altering the existing infrastructure, despite having to sacrificing a bit of the positioning accuracy, as well as the surveying labour cost, and the maintenance overhead of such database.

5.7 | Overall remarks and open research questions

Overall, VL-based system closely resembles the way human navigates. As a promising indicator, it offers some of the highest positioning accuracy at the time of writing, up to centimetre level with custom made LEDs, whereas infrastructure-free systems using just the smartphones camera may still offer 1 m to 2 m accuracy. Nevertheless, there are still challenges from the sensor design (e.g. smartphones camera has slow frame rate, narrow field of view, etc.), the lack of depth information (i.e. most current smartphones are not equipped with ToF sensor). With more and more smartphones come equipped with ToF sensor, VL systems may play a major part in future mainstream indoor positioning service.

Research wise, potential academic researchers may be interested in the following open questions:

- *Lighting conditions*. As the smartphones camera completely rely on natural visible light to operate, over-exposure (i.e. too much lights) or under-exposure (i.e. too little light) will impact the captured optical information.
- *Viewing angle*. Most smartphone cameras have narrow viewing angle. In addition, the user tends to tilt the phone downwards while viewing the screen, which decreases the angle even further.

6 | INERTIAL BASED SYSTEMS

While previously discussed electromagnetic and light-based systems rely on external references (e.g. radio signals, light, etc.) to function, inertial-based systems use the phone motion to estimate the change of position in relation to the starting point. Those sensors include the accelerometer measuring the acceleration, and the gyroscope measuring the rotation. In the smartphones, those sensors are coupled together in one inertial measurement unit (IMU), often with the magnetometer described in Section 7.4. The rationale is that although magnetometer does not directly measure motions, its function in detecting the Earth's magnetic North (e.g. as a compass) may be used in conjunction with accelerometer and gyroscope to determine the absolute heading. Nevertheless, the use of magnetometer for inertial tracking in this context could be restricted by the strong anomalies of the magnetic field within the building [54].

Nevertheless, as with other sensor categories, inertial sensors were not designed with indoor positioning in mind. Hence, they face the following challenges.

- *Unconstrained phone orientation*. As the sensor is fixed inside the smartphones, its 3-dimensional (3D) coordinates are aligned with the phone's body frame (see Figure 12). Hence, we observe different measures with different phone's orientations, in the same spot.
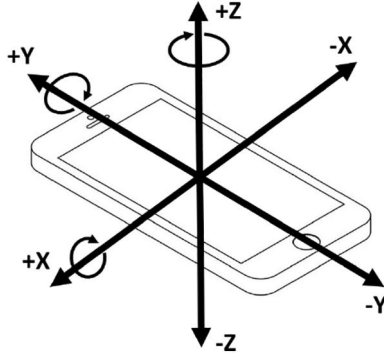
FIGURE 12 The inertial sensor's co-ordinate is aligned with the phone's body frame

- *Uncontrolled body placement*. We have no control on how the user holds their device. As the users move, the placement of the smartphone on their body has a major impact on the sensor measures.
- *Unrelated measures*. For positioning purpose, we are only interested in true movement-related force. However, other gestures such as arm, body movements are recorded too.

6.1 | Accelerometer

The accelerometer is designed to measure acceleration forces, which are the changing rate of linear velocity. Such forces may be static such as the Earth's continuous gravitational pull, or dynamic such as the phone vibrations and human movements. The measuring unit is in metre per second squared (m/s^2). The available information is a 3D vector corresponding to the acceleration force being applied at each Cartesian axis of (x, y, z).

On the smartphone, the accelerometer was originally intended to detect which way the phone is pointing (i.e. based on the Earth's static gravitational pull whose strength and direction are known and constant), in order to adapt the screen into portrait or landscape mode accordingly. Modern smartphones are capable of high sampling rate of up to 500 Hz to detect the slightest change in movements.

6.2 | Gyroscope

The gyroscope is designed to measure rotary motion, which is the changing rate of the angular rotational velocity. The measuring unit is in degree per second (°/s). The available information is a 3D vector corresponding to the rate of rotation around the three Cartesian axes (x, y, z).

On the smartphone, gyroscope was designed to measure how much the phone is tilting. On a quick thought, this is what an accelerometer could achieve. However, the critical difference is that an accelerometer may only determine the phone's orientation when it is static, by analysing the Earth's gravitational pull [55]. When the phone is moving, the true acceleration is mixed with the Earth's gravity, which cannot be easily separated.

On the other hand, the gyroscope can measure precisely the phone's rotary angle in 3D. By combining accelerometer and gyroscope, the phone can better sense its position for many motion-sensing applications such as Virtual Reality headset and image stabilisation. Similar to accelerometer, modern smartphones are capable of high gyroscope sampling rate.

6.3 | Positioning methods

Having introduced the inertial sensors, we will now overview the most popular positioning methods. In principal, there are two main approaches for inertial tracking, that are strap-down and step-and-heading.

However, the shared challenge for both approaches, as discussed earlier, is that while the sensor measures are aligned to the phone body frame, the phone itself is not fixed to any particular axis, and is likely to be shifted around in the hand or pocket as the user walks. Hence, to work out the true acceleration, we use the gyroscope to track the orientation of the phone, and in doing so, being able to maintain a consistent global reference frame for the sensor measures.

6.3.1 | Tracking the phone orientation

The steps to track the phone orientation are as follows. First, we need to convert the sensor's local body frame, which changes as the phone is rotated, into a global fixed frame. Given the angular rotational velocity provided by the gyroscope at time point t: $g(t) = (g_x(t), g_y(t), g_z(t))$, we specify the following 3×3 rotation matrices O_x, O_y, O_z, where each column represents a vector along the global frame axis.

$$
\begin{aligned}
O_x(t) &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & -cos(g_x(t)) & sin(g_x(t)) \\ 0 & sin(g_x(t)) & cos(g_x(t)) \end{bmatrix} \\
O_y(t) &= \begin{bmatrix} cos(g_y(t)) & 0 & sin(g_y(t)) \\ 0 & 1 & 0 \\ -sin(g_y(t)) & 0 & cos(g_y(t)) \end{bmatrix} \\
O_z(t) &= \begin{bmatrix} cos(g_z(t)) & sin(g_z(t)) & 0 \\ -sin(g_z(t)) & cos(g_z(t)) & 0 \\ 0 & 0 & 1 \end{bmatrix}
\end{aligned} \tag{14}
$$

Then, the new sensor frame is computed as $O_T = O_x O_y O_z$. To transform the new sensor measures to the global frame, we multiply them with O_T [56,57].

Once the sensor measures are in the same consistent global frame, we can safely subtract the Earth's gravity vector $g = (0, 0, 9.81)$, which is always pointing down on the z-axis, from the accelerometer vector $a(t) = (a_x(t), a_y(t), a_z(t))$, to reveal the true acceleration of the phone at time point t, as follows:

$$a(t)_{noG} = a(t) - g \qquad (15)$$

The above true acceleration vector of the phone can then be used with the strap-down and step-and-heading approach below.

6.3.2 | Strap-down approach

With strap-down approach, the phone positions are derived by continuously estimating the velocity based on the measure provided at fixed periods by the accelerometer.

Without loss of generality, the velocity $v(t)$ of the moving phone at time point t can then be computed as follows:

$$v(t) = v(0) + \int_0^t a(t)_{noG} dt, \qquad (16)$$

with $v(0) = 0$ is the initial velocity when the device is static.

The above velocity is then integrated again to produce the phone's position.

$$p(t) = p(0) + \int_0^t v(t) dt, \qquad (17)$$

with $p(0) = 0$ is the initial position.

6.3.3 | Step-and-heading approach

With step-and-heading (SaH) approach, the phone positions are estimated via the user's continuous walking steps and strides. Therefore, this approach is pedestrian specific. In general, SaH contains three steps:

Step 1 *Extracting step-related measures.* As the sensor measures contain the acceleration and rotation of the phone in general, the first step is identifying which segments of the sequence represent the walking steps (i.e. to count how many steps the user has walked). There are two options for this task.

- Stance detection: This approach identifies the sensor segment in which the user foot is planted on the floor (i.e. to be counted as one step). This is done by simply checking if the sensor measure is static (i.e. below a certain threshold) which implies the user is not moving at that moment.
- Step cycle detection: This approach looks for repeated patterns within the sensor sequence, assuming that the user has a relatively consistent gait throughout.

Step 2 *Estimating step length.* Once the relevant step related sensor sequence has been extracted above, each segment will be counted as one walking step. The length of each step can simply be defined as a constant (e.g. 70 cm). This is a relatively practical assumption for natural walking pace [58]. However, the human stride may vary significantly by up to 50% when the walking speed changes [59].

Step 3 *Estimating heading.* The user's heading is then inferred in a similar manner as in strap-down system, by integrating the gyroscope measure to obtain the heading change estimation. However, the advantage is that there is no need for a second integration of the accelerometer measure to work out the travelling distance as we may just use the step length above. Therefore, the accumulated error grows linearly, rather than cubically in strap-down approach.

Without loss of generality, the position of the moving phone at time point t can be computed as follows:

$$p(t) = p(t - 1) + step * cos(\theta(t)) \qquad (18)$$

where *step* is the step length and θ is the heading.

Figure 13 visually summarises the processes in an inertial tracking system.

6.3.4 | Incorporating floor-plan information

With inertial tracking systems, the user position is estimated sequentially, in relation to the starting point (whereas wireless signal-based or VL-based systems can estimate the user position independently at any moment without reference to previous estimations). Therefore, any error happening during the process (which unfortunately is inevitable at both hardware and software levels, because of the sensor orientation and the integration in the algorithm) will accumulate rapidly, the longer the user trajectory is.

As such, inertial systems need other source of information to correct these position drifts. One such information is the floor-plan of the tracking zone, which details the geometry shape of the walls, rooms, corridors. The working principle is that the transition between two consecutive positioning estimations must not violate the geometric constraints imposed by the floor-plan (i.e. the user cannot walk through walls, or jump to the other side of the corridor). There are two popular implementations of such information in the literature.

- *Refining the user positions.* If the starting point on the map is known before-hand, the system just actively monitors the position estimation at each step, making sure it is within the boundaries on the map.
- *Constraining the walking trajectory.* If the starting point is unknown, the system builds up the shape of the trajectory along the way. At some point, the path is long and distinctive enough to be matched against the floor-plan to provide an estimation of the user's whereabouts.

6.4 | Performance review

Having introduced the inertial-based techniques, we are now in a good position to review their performances on real world systems (see Table 5).
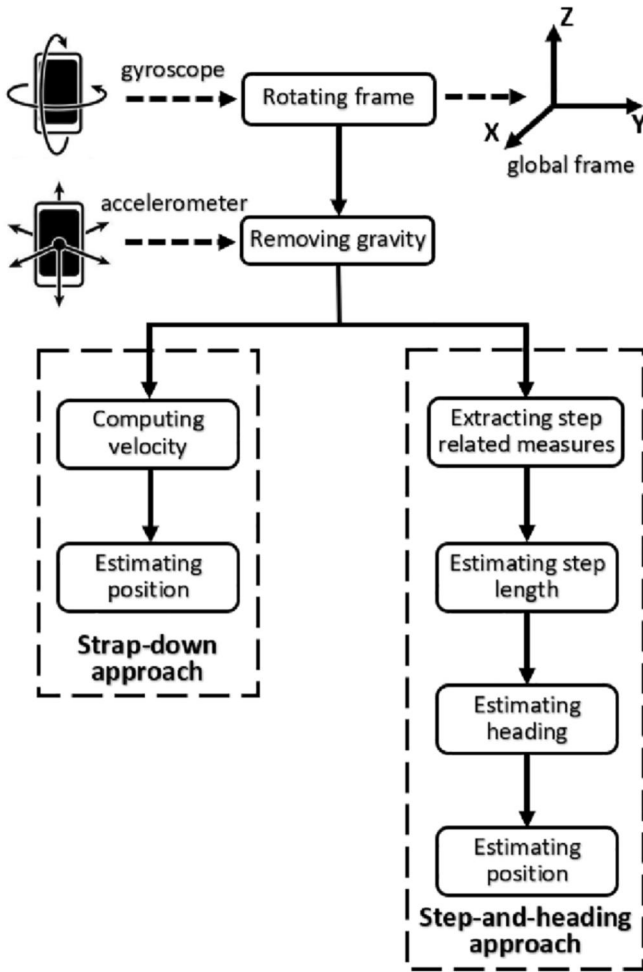
FIGURE 13 The two main approaches for inertial-based system

equipped with inertial sensors to support the basic functions of adapting the phone screen's orientation. Accessibility wise, from the indoor positioning service's perspective, this group of sensors can be accessed at any given time by the app, without any user permission, to deliver a smooth user experience, yet some may argue about its potential security risk [54]. Power wise, inertial sensors belong to the group of so-called low power sensors, which consume so little power that Android allows them to be always on.

Positioning accuracy wise, inertial systems achieve around 1 m on average, which is great considering its advantages above. However, such accuracy quickly degrades the longer the user walks. Hence, without ground-truth data to correct the sensor and estimation error, such systems may not be usable in the long run.

Research wise, potential academic researchers may be interested in the following open questions:

- *How to detect various step lengths?* Most inertial systems assume a constant step length throughout the user's journey, which may not be true for different walking paces. As stride length is an important parameter in the positioning estimation equation, it should be adjusted in real time.
- *How to robustly estimate the sensor heading?* The user may hold the phone in hand, shirt pocket, trouser pocket, bag etc. which impacts the sensor measures. The magnetometer which was designed to track the phone's heading, does not work reliably indoors because of the high magnetic anomalies.
- *How to address sensor drifting?* This is perhaps one of the biggest challenges. In addition, the phone sensors are noisy which amplifies the estimation errors.

7 | OTHER SYSTEMS

This section overviews the sensors that do not fit directly in the above three categories. Unfortunately, at the time of writing, most of these sensors are not explicitly useful for indoor positioning, but we include them here for completeness.

7.1 | Microphone

The microphone is an acoustic sensor that converts sound waves into electrical signals. In contrast to the previously discussed electromagnetic waves that can propagate seamlessly through space, sound waves are mechanical waves that require a material medium such as air or water to propagate from one place to another.

On the smartphones, the microphone was designed for human speech recording. Hence, its sensitivity was factory-adjusted to specifically aim at human audible voice ranging from 20 Hz to 20 kHz. However, some microphones could pick up inaudible sounds beyond this range [66,67]. The available information is the raw sound data, and the signal amplitude. The later information could be useful to determine

Technique wise, strap-down approach performed slightly worse than step-and-heading one [60,61], as it has to perform the double integration, as explained earlier. Since the phone is not fixed to the user body, any small errors in estimating the acceleration and rotation will translate into progressively bigger error after each integration. Generally speaking, the longer pure inertial-based systems are executed, the larger the error will accrue, and the further the estimated position will deviate from the true location, which may be as much as 17 m after just 550 steps [62]. To tackle this challenge, Zhao et al. [17] infers the magnetometer data to reduce the heading drift, whereas Solin et al. [60] exploits the occasional user stationary states to reset the velocity. Nevertheless, the indoor magnetic field may be heavily distorted indoors, and the user may keep walking for long period of time [63].

6.5 | Overall remarks and open research questions

Inertial-based approach's appeal lies in the scalability as the system is self-contained and requires no additional infrastructure. Availability wise, almost every single smartphone comes

TABLE 5 Comparison of notable inertial-based system performance

Authors	Sensor	Positioning technique	Testbed	Accuracy	Notes
Kang et al. [61]	Accelerometer, magnetometer, gyroscope	Step-and-heading	85 m walking path, 1 min walk	1.23 m mean error	The authors propose to select the most likelihood time between two consecutive steps where the heading direction may change. They tested their system on a smaller 75 m walking path, with slightly better accuracy.
Zhao et al. [64]	Accelerometer, magnetometer, gyroscope	Step-and-heading	858 m walking path, 664 steps	1.57 m mean error (± 0.95 m)	Gradient descent algorithm was implemented to reduce the heading drift using the magnetometer data which was distorted by hard-iron.
Solin et al. [60]	Accelerometer, gyroscope	Strap-down	40 m path, 2 min walk	1.8 m mean error	An extended Kalman filter was applied to fuse the accelerometer and gyroscope measures. The model is constrained by zero-velocity updates when the user is stationary.
Qian et al. [65]	Accelerometer, magnetometer, gyroscope	Step-and-heading	18×12m lab, 60 m walking path	1.71 m, 95% chance, phone in pocket	The floor-plan was used to constrain the inertial drifting. The positioning accuracy was higher at 1 m, 95% chance when the phone was in hand.

the distance between the sound transmitter to the receiver, as sound waves gradually lose their energy in the air, in the same manner as electromagnetic waves.

Modern smartphones may have two microphones, one mouthpiece located at the bottom of the phone, where it is closest to the human mouth for the loudest voice recording, and a back microphone located in the back of the phone to capture the ambient sound. The combination of two audio sources will help with noise reduction and improving sound quality.

The clear advantage of microphone is that it is a ubiquitous sensor on every single smartphone to facilitate the essential function of audio communication. Nevertheless, it faces the following challenges.

- *Audio level*. High frequency noise can be easily blocked by covering the microphone (e.g. when it is inside pocket or bag). Similarly, the clarity of the audio recording can be impacted by changing the sensor orientation.
- *Audio distortion*. Microphone has a volume level upper-bound which it may handle. When the incoming sound is too loud, the captured recording may be distorted.
- *Low temporal variation*. The sound within an indoor space tends to vary significantly over time by infrequent noises (e.g. people chatting occasionally, machine noise during daytime but not night-time). This has a major impact on systems mapping sound to location (e.g. fingerprinting).

Overall, microphone offers a means to record sound, which may be useful for some indoor navigation systems.

7.2 | Barometer

The barometer is designed to measure air pressure (i.e. the weight of the air exerted by the overhead atmosphere onto a surface or object). The sensor was originally meant for weather forecasting, as air pressure is caused by the amount of water within the air. The higher the reading is, the better the weather will be. In contrast, the lower the reading is, the more likely it will rain. The measuring unit is Pascal (Pa).

On the smartphone, however, barometer is purposely used to improve the GNSS accuracy with the added altitude information. The observation is that the atmospheric pressure decreases as altitude increases. There exist many models to map altitude to pressure [68,69]. However, as barometer was not originally designed for such purpose, it faces the following challenges.

- *Closed environment*. Barometer works best in clear, open air. The air indoors circulate in a closed area, where air-conditioner, fans may impact the reading.
- *Low sensitivity*. The sensor measure does not vary significantly with respect to small altitude changes. It was reported that the pressure drops just 11 Pa for each metre [69]. Therefore, this measure may be lost among other sensor noises.

Overall, barometer is useful to distinguish users on different floors in buildings with high ceiling, rather than determining fine-grained positions on the same floor level. It adds the extra z-dimension into the positioning information to complement other systems.

7.3 | Proximity

The proximity sensor is designed to detect objects within its vicinity without any physical contacts. To do so, it emits an electromagnetic beam (e.g. invisible infrared LED) and senses any light that bounces back. The more light being reflected back, the closer the object should be. The active working range is fairly short at about 6 cm, which is slightly shorter than NFC's.

On the smartphones, the proximity sensor is situated by the front camera and the ear speaker (see Figure 2), with the sole purpose of automatically turning off the screen while the user is holding the phone by their ears (by detecting the side of the user's face) to avoid unintended touches. Although the sensor is capable of estimating the distance to the nearest object in centimetre, some smartphones' firmware restricts the output in simple binary form of being near or far.

At present, because of the inconvenient placement and the restricted output, this sensor does not appear to be useful for indoor positioning. Consequently, there has been no published literature that employs such technology. Nevertheless, we include it here for completeness.

7.4 | Magnetometer

The magnetometer is designed to measure the strength and direction of the static magnetic fields (e.g. the Earth's magnetism, the magnet bar). These static fields do not change over time nor have a frequency (i.e. 0 Hz). Whereas, other electromagnetic signals discussed so far (e.g. WiFi, Bluetooth, GNSS, FM and Cellular) are dynamic, in which their waves with different frequencies travel in the air.

On the smartphones, this sensor was intended to identify the true North for outdoor navigation, made popular by the compass app. It may also be used as ferrous metal detector. Measurement-wise, the available information are the magnetic field strength, in microTesla unit (μT), along the three axis of the phone. Modern device has a sampling rate of up to 100 Hz. The advantage of magnetic-based system is that magnetism exists everywhere on Earth, and requires no additional infrastructure.

However, magnetometer-based system faces the following challenges.

- *Phone placement in 3D.* Since the sensor's frame aligns to the phone's body, any change to the phone's orientation in the 3D space varies the three sensor's measures, even in the same position.
- *Magnetic disturbance.* The compass does not work accurately indoors, because of the building materials (e.g. metal bars, steel rebars, reinforced concrete) which greatly distort the Earth's magnetic field.

Overall, magnetometer provides a means to capture the ubiquitous magnetic field that exists in all buildings, and may be valuable for some indoor positioning systems.

7.5 | Positioning method

Besides fingerprinting, another method that is applicable to acoustic sound based system is the time difference of arrival (TDoA). The motivation for this method is that acoustic sound wave travels much slowly (i.e. 343 m/s) than the WiFi and Bluetooth counterparts (i.e. the speed of light at 300,000 km/s).

Hence, it is easier for mobile devices to compute the distance between them by timing the total time it takes for a chirp to travel, given that the system clock from both smartphones are perfectly in sync, as follows [73,74]:

$$distance\ =\ speed\ of\ sound\ *\ elapsed\ time \qquad (19)$$

where speed of sound is a constant.

Nevertheless, the challenges for this approach are that it is difficult to perfectly synchronise the internal clock on various smartphones. In addition, the existing Android APIs contain certain latency, which delays the transmitting time on the sender, hence introducing significant error in the distance estimation.

7.6 | Performance review

This section only reviews the performance of microphone in real world systems, as this is the only sensor in this category, capable of powering the entire positioning system alone (see Table 6). There are two main types of microphone-based systems. The first type generates a fingerprinting database using passive ambient sound. The second type deploys additional speakers around the building with different noise levels and patterns, then record a fingerprinting database in a similar manner.

Positioning accuracy wise, systems that made use of dedicated speakers achieved much higher accuracy at just 50 cm error, compared to those using natural ambient sound at room-level [70,72]. This result is not surprising given most indoor spaces have indistinguishable ambient noise. Rossi et al. proposed an interesting approach where the in-built smartphone speaker is used to broadcast a short acoustic noise, which is then captured by the microphone on the same device, hence avoiding the use of additional infrastructure [71]. Their assumption is that the acoustic impulse response is different for each location, to generate a fingerprinting database. Their results demonstrated 98% room-level accuracy, 51.3% within-room accuracy, tested in 20 different rooms.

Overall, most sound-based positioning systems offer around room-level accuracy, although those making use of dedicated speakers may offer sub-metre accuracy. Nevertheless, there are concerns regarding the sustainability of such accuracy, as indoor sound's temporal variation is high (e.g. an office is more noisy during daytime than night-time).

8 | COMPARISON OF TECHNOLOGIES

Having reviewed individual sensor system, the authors may now compare their performances against each other, in two different settings, the general setting and the competition setting. In the general setting, they will use the accuracy reported in the literature. The benefit of this setting is that they have access to a large pool of systems with detailed results and

TABLE 6 Comparison of notable sound-based system performance

Authors	Sensor	Positioning technique	Testbed	Accuracy	Notes
Tarzia et al. [70]	Microphone	Fingerprinting	43 rooms, 2 training positions per room	60% room-level accuracy	Natural ambient noise was recorded, then converted into power spectrogram to generate a training database. Nearest neighbour using Euclidean distance metric was employed for location matching.
Rossi et al. [71]	Microphone, speaker	Fingerprinting	1 training position every 9 m^2, 67 in total across 20 rooms	98% room-level accuracy, 51.3% within-room accuracy	The smartphone speaker was used to broadcast noise at different locations to generate a training database. SVM was employed for location matching.
Rishabh et al. [72]	Microphone, speaker	Trilateration	6 m × 5 m × 2.6 m room, 6 ceiling mounted speakers	50 cm, 90% chance	Standard PC speakers were deployed to play in-audible low frequency noises. A sound propagation model was employed to estimate the distance from the smartphone microphone to the speakers.

implementations, despite the apparent lack of ground-truth verification. In the competition setting, they will overview the performance reported in the Microsoft Indoor Localisation Competition—ISPN, from 2014 to 2018; and the Indoor Positioning and Indoor Navigation (IPIN) Indoor Localisation Competition, from 2015 to 2019 (the first three competitions employed a scoring metric instead of the standardised metre unit, to rank the competitors). The benefit of this competition setting is that all participating systems were evaluated on the same testbed, despite the lack of details in the implementation (i.e. systems from the industry often do not reveal their design).

8.1 | Comparison in the general setting

There are several metrics to rank different indoor positioning technologies. As the users may prioritise different criteria depending on their applications, the authors will review most criteria here, and let our readers decide on which one is suitable for them.

In particular, the criteria for assessment will be accuracy, latency, cost, maintenance, and scalability. Security, despite being an important factor, is not part of this review, as there are many external factors that may impact the system. Table 7 compares the sensor performances using the above metrics.

8.1.1 | Accuracy

Accuracy is arguably the most considered factor for a positioning system. This metric will decide if the system is capable of achieving fine-grained tracking with precise positions, or just coarse-grained proximity tracking in a wide area. It reports the difference between the estimated position by the system and the true position. Some systems average these results over multiple testing locations, others report the variation of the above positioning accuracy (e.g. 2 m accuracy, 95% of the

TABLE 7 System performance under different metrics

Sensor type	Accuracy	Latency	Cost	Scalability
Barometer	Floor-level	High	Low	High
Bluetooth	Room-level	High	Medium	Low
Camera	High	Low	Changing	High
Cellular	Town-level	High	Low	High
GNSS	Low	High	Low	High
Inertial[a]	Changing	Low	Low	High
Magnetometer	Medium	Low	Low	Low
Microphone	Low	Low	Low	Low
WiFi	Medium	High	Low	Low

[a]Inertial sensors include the accelerometer and gyroscope.

time). Normally, the cumulative distribution function (CDF) is employed to report the system's accuracy.

The results in Sections 4.8, 5.6, 6.4 and 7.6 suggested that VL-based systems achieved some of the highest positioning accuracy at centimetre level, using custom LEDs. This is followed by WiFi, Bluetooth, and Camera systems (using indoor images) at around 2 m to 3 m accuracy. Inertial systems start off rather accurately at sub-metre level, but quickly degrade exponentially the longer the travelling trajectory is. Other systems such as sound-based, magnetometer-based, and FM-based, typically achieved room-level accuracy.

8.1.2 | Latency

Latency measures the delay for which the information is provided. This metric is critical for tracking fast moving users. It is a combination of the hardware latency (i.e. how frequent the sensor produces a measure), and the software latency (i.e. the algorithm executing time).

Sensor level wise, inertial sensors have the highest sampling rates at more than 100 Hz (100 samples per second) on most modern smartphones (see Table 2). This design was intended to capture fast motions. At the other end of the spectrum were WiFi and Bluetooth, with 1.5 Hz and 0.25 Hz, respectively.

Algorithm level wise, systems that rely on a database may impact the real time positioning speed, if the algorithm requires a full scan of the entire database to estimate the user position (e.g. K-nearest neighbours). A model based approach (e.g. SVM) would accelerate systems involving a large training database.

8.1.3 | Cost

Cost is the total expense to deploy and maintain the positioning system. In general, inertial-based approach offers the most cost effective solution, as it needs no additional hardware or software maintenance. The entire system may run locally on the user's smartphone. Systems that make use of ubiquitous infrastructure (e.g. WiFi, magnetic field, etc.) have no infrastructure cost, but may require intensive surveying effort to generate the training database, not least when the tracking zone is large, as well as the maintenance overhead to update such database over time.

The general trade-off is that the higher the system cost, the more accurate the system may be, and vice-versa (see Figure 14).

8.1.4 | Scalability

This metric is the system's capability to serve a huge number of users simultaneously, with plenty of room for future expansion. This is the most essential metric for commercial solutions, but is perhaps one of the most under-considered metrics for indoor positioning systems.
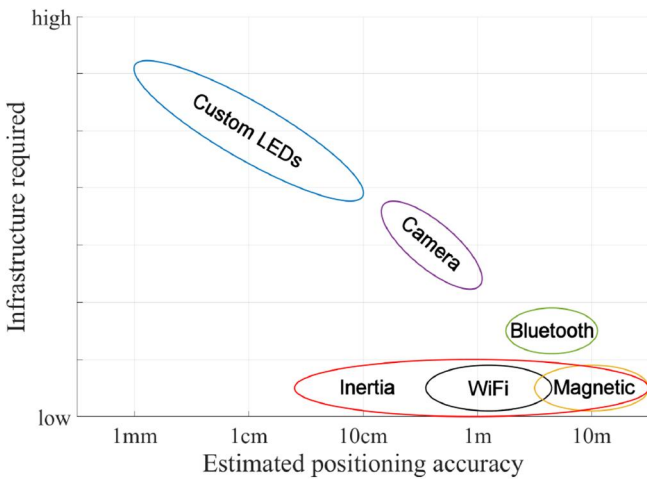

FIGURE 14 A rough correlation between the amount of infrastructure and the estimated accuracy of indoor positioning technologies. The x-axis is in log scale

Generally speaking, scalable systems should function in a decentralised fashion, where the user's smartphones perform most of the work. For example, systems that operate locally on the smartphones (e.g. interial-based, custom LEDs) may easily scaled to the vast number of users. On the other hand, systems using a central server (e.g. fingerprinting-based) may struggle to serve many users simultaneously, including the bandwidth cost for communication between the users and the server [75].

8.2 | Comparison in the competition setting

Having been introduced since 2011 with the IPIN Indoor Localisation competition and since 2014 with the Microsoft Indoor Localisation competition, these competitions brought together a large number of academic institutions and industrial companies. Beside the competitive nature, they offer a common test venue, spread over multiple buildings and floors, with no advantageous inside information to fine-tune the participants' system. Therefore, these competitions offered a fair performance comparison among positioning technologies. It is worth noting that since there were different tracks, our focus will be on the smartphone-based systems (i.e. other proprietary technologies such as LIDAR, laser scanning, UWB, etc. will not be reviewed, despite constantly being the top performing systems).

8.2.1 | The Microsoft indoor localisation competition

The aim of the Microsoft sponsored competition is to enable a wide-range of positioning technologies, tested in a small-to-medium venue, where just a simple mean error metric was employed to rank the competitors. In all 5 years of the competition, it appeared that WiFi, BLE, and IMU are the most popular choices, among all smartphone-based sensors (see Figure 15).

Interestingly, for single technology-based system, IMU systems achieved the highest accuracy at 1.37 m mean positioning error, as well as the highest overall accuracy at 2.45 m. Unsurprisingly, most of the top performing competitors incorporated IMU in their systems, with the most accurate system achieved 1.23 m error, involving a combination of WiFi, magnetic fingerprinting, and IMU (see Figure 16).

8.2.2 | The IPIN indoor localisation competition

The aim of the IPIN competition is to provide a large, multi-building, multi-floor testbed, specifically aiming at certain positioning technologies (e.g. smartphone-based, foot-mounted-based). The ranking metric was the third quartile of the positioning error. Similar to the Microsoft counterpart, the IPIN Indoor Localisation competition also attracted a large number of WiFi, BLE, and IMU-based systems. Interestingly, it appeared that the huge testbed of the IPIN competition did
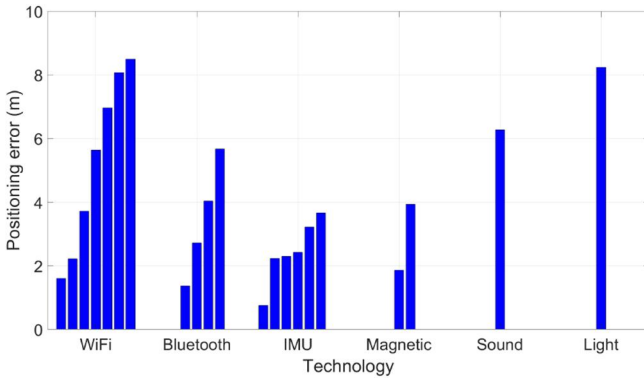
FIGURE 15 Comparison of the performance accuracy of single technology-based system participated in the Microsoft Indoor Localisation competition between 2014 and 2018
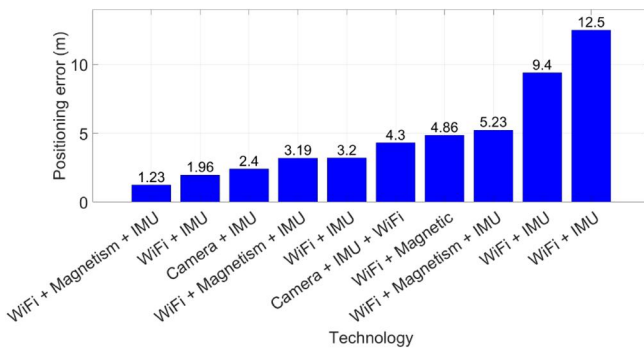

FIGURE 16 Comparison of the performance accuracy of hybrid systems participated in the Microsoft Indoor Localisation competition between 2014 and 2018

affect the accuracy of magnetism-based systems, where some of them struggled with more than 15 m positioning error. At the other end, BLE-based systems appeared to perform well in this setting (see Figure 17).

Overall, average positioning error wise, the top systems from both competitions achieved similar result at round 4–5 m, with the IPIN ones scoring a bit higher because of their larger testbed. Single technology wise, IMU-based systems managed the highest accuracy at around 2.45 m, in the Microsoft competition; whereas WiFi fingerprinting-based systems were the most accurate one at around 10.22 m, in the IPIN counterpart. Hybrid system wise, a combination of WiFi, Magnetism, and IMU produced the most accurate system for the Microsoft competition, at just 1.23 m error; whereas the combination of WiFi and BLE was the most accurate one for the IPIN competition at 3.48 m error.

9 | HYBRID SYSTEMS

Having discussed the strengths and weaknesses of individual sensor in previous sections, the overall consensus is that there is no stand-alone best option for all indoor scenarios. As each sensor excels in certain condition, it is not surprising that most
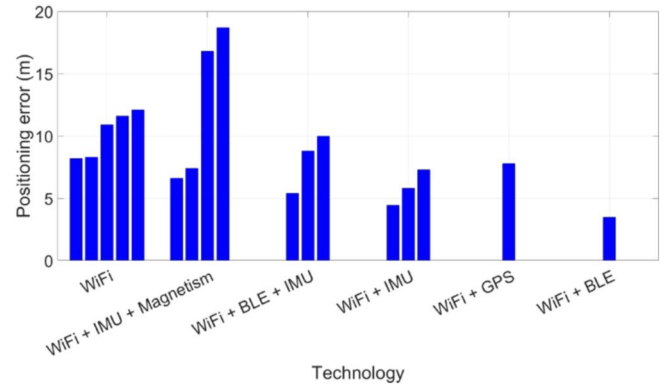

FIGURE 17 Comparison of the performance accuracy of positioning systems participated in the IPIN Indoor Localisation competition between 2015 and 2019

effort were invested to combine such sensors together. This section reviews hybrid systems that involve two or more of the sensors described above, and the fusion techniques behind.

9.1 | Sensor fusion techniques

In indoor positioning, sensor fusion is the process of combining data from several sensors to improve the positioning accuracy. There are three different categories in sensor fusion [76,77].

- *Complementary category*. With this category, each sensor is responsible for a separate, disjunctive part of the environment, and their measures will later be combined to provide a fuller knowledge of the indoor space. For example, two cameras were set up to cover different indoor areas.
- *Co-operative category*. This category uses the measures provided by different sensors to derive new information that was not available before. For example, the combination of camera and time-of-flight sensor provides a new 3D view of the indoor space.
- *Competitive category*. With this category, different sensors of the same type independently provide measures about the same property. For example, two cameras provide two images of the same object. However, one image may be blur due to poorer lens. Hence, the fusion of both images will result in a clearer result with less noise.

It is worth emphasising that it is possible to employ more than one approach in the same system, for example, a positioning system using a network of Bluetooth and WiFi beacons deployed at different locations in the building. Each beacon covers a distinct area in a complementary fashion, whereas WiFi and Bluetooth beacons in the same location provide similar distance measure RSS to the smartphone in a competitive fashion. Overall, most smartphones-based systems would fall into the co-operative category, because all sensors are housed together within the device. They mostly observe different information of the same indoor space.
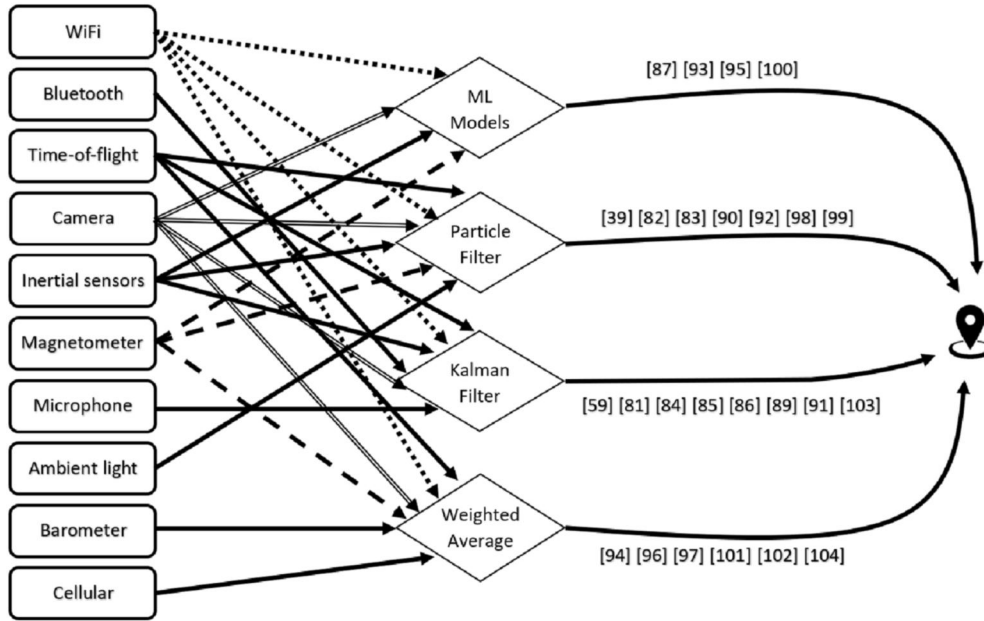
FIGURE 18 Combination of smartphones sensors with respect to the data fusion methods. Different styles of arrows are used to distinguish among sensor types and how they combine

Once the sensors' approach has been decided, there are two settings to process their measures as follows.

- *Centralised setting*. In this setting, the measures from all sensors must be available simultaneously during the fusion process.
- *Decentralised setting*. In this setting, a model is derived for each sensor. When the fusion happens, each sensor model is processed sequentially.

With the smartphones, all sensors can be queried at any given moment. Although they have different sampling rates, faster sensor can be adjusted to match the sampling rate of slower ones. Hence, both of the above settings can be employed for data fusion.

Once the approach and setting have been decided, the fusion algorithms will be applied onto the sensor measures. Within the indoor positioning literature, three popular data fusion algorithms included Particle, Kalman Filter, ML-based (e.g. Hidden Markov Model, Gaussian Mixture Model, Bayesian inference, etc.) [78], and weighted average. We briefly discuss their core concepts below.

- *Filtering-based algorithm*. Filtering algorithms are those that recursively refine a set of positioning estimations over time, given a sequence of sensors' measure. The Kalman Filter, for example, applies linear projections, whereas Particle Filter uses sequential Monte Carlo method [79–81]. Each state normally includes the positioning label, the weight representing the probability of it being the true position, and other set of information related to the sensors being used.

- *Machine learning-based model*. ML models may be used to represent the distribution of each sensor's measure, which can later be fused using a mixture model (e.g. Gaussian Mixture Model) [82,83]. This is mostly used in the decentralised sensor setting described earlier.
- *Weighted average*. With this approach, each sensor produces its own positioning estimation independently. The fusion simply takes the average of the each sensor's results with higher weights awarded to more reliable sensors.

Figure 18 visualises the combination of smartphones sensors with respect to different data fusion methods in this review. Two filtering-based algorithms, Particle Filter and the Kalman Filter, are separated in the figure for better highlights.

9.2 | Performance review

With 18 smartphones sensors reviewed above, even after leaving out non-relevant ones to indoor positioning, there are still too many possible permutations of hybrid systems to review. Hence, we mostly focus on combinations with the potential to improve the overall accuracy, where sensors complement for the weaknesses of each other. In particular, some potential matchings at sensor category level are:

- *Inertial and electromagnetic sensors*. Inertial-based systems start off very accurately (i.e. within the first few steps), but cannot sustain such performance over long distance as sensor drifting happens. WiFi or Bluetooth beacons, on the other hand, could be used to correct such drifts [84–87].

T A B L E 8 Comparison of notable hybrid system performance

Authors	Sensor	Fusion technique	Testbed	Accuracy	Notes
Zhu et al. [86]	Inertial sensors, WiFi	Particle filter	362.6 m^2, 4 WiFi APs, 30 training locations	1.5 m, 90% chance	WiFi fingerprinting and the building map were employed to correct the drifting error from inertial sensors. The proposed method achieved 1.2 m mean error, compared to 2.1 m using inertial sensors and 2.8 m using WiFi.
Maghdid et al. [87]	Inertial sensors, WiFi	Extended Kalman Filter	48 × 38 m corridor, 7 WiFi APs	4 m, 90% chance	Inertial tracking alone over 172 m travelling distance drifts to over 40 m positioning error.
Li et al. [84]	Inertial sensors, Bluetooth	Extended Kalman Filter	20 Bluetooth beacons over a university corridor	2.26 m mean error	Bluetooth beacons were used for proximity tracking without a training database to correct the inertial sensors' drifting. The positioning results with only 15 and 10 beacons were degraded by 28.9% and 43.8%, respectively.
Putta et al. [85]	Inertial sensors, magnetometer	Particle filter	9 × 16 m area, 3 walking trajectories of 85 m each	1.5 m, 90% chance	Gradient descent algorithm was employed to correct the inaccuracies in the user heading estimates due to magnetic perturbations.
Al Delail et al. [88]	Inertial sensors, camera	Extended Kalman Filter	309.4 m walking trajectory	1 m mean error	Principal component analysis was employed to identify the direction of motion.
Xu et al. [93]	Inertial sensors, ambient light	Particle Filter	3 buildings with 120 luminaries	50 cm, 90% chance	Illumination peak detection algorithm was used to distinguish various light intensity amongst locations. The information are then fused with inertial sensors measures.
Yang et al. [94]	Inertial sensors, microphone	Kalman Filter	8 × 7 m room, 6 ceiling mounted sound receivers	50 cm mean error	Acoustic pulses emitted from the smartphone every 0.3 s are captured by the sound receivers to determine the position.
Shu et al. [95]	WiFi, magnetometer	Particle Filter	47 × 85 m office space	90 cm, 90% chance	The magnetic and WiFi fingerprinting databases were combined under a particle filter for each scan. The result with only the magnetic field was 1.5 m, 90% chance, and 2.5 m, 90% chance with only the WiFi database.
Shao et al. [96]	WiFi, magnetometer	Convolutional neural network	60 × 40 m office space	1 m, 90% chance	The WiFi and magnetic signals were converted into a fingerprint image to represent the features of each location. These images are then input into a deep CNN to predict real-time location estimates.
Baniukevic et al. [97]	WiFi, Bluetooth	Weighted average	50 m^2, more than 30 WiFi APs, 12 training points	1.75 m mean error	Pure WiFi positioning achieved 3.15 m mean error. Pure Bluetooth positioning achieved 7.57 m mean error. The positioning result in a large shopping centre with less stable signals was about 7 m mean error.
Liu et al. [98]	WiFi, microphone	Graph mapping	12 × 11m office, 71 training locations, 14 WiFi APs	1.2 m, 90% chance	When a phone needs to improve its positioning accuracy, it sends a special audio signal to recruit other nearby phones. Acoustic ranging estimates amongst phones were used alongside the WiFi fingerprint map.
Bisio et al. [99]	WiFi, barometer	Weighted average	4 floors, 150 m^2 and 28 WiFi APs per floor	1.5 m mean error	Barometer measure was much more effective when fewer APs were detected. WiFi positioning only achieved 4.1 m mean error. Adding barometer would increase the result to 2.5 m mean error.
Xu et al. [100]	WiFi, camera	Weighted average	Shopping mall, 1.8 m space in training points	2.2 m, 90% chance	Geometric constraints were extracted from images to reduce WiFi fingerprint ambiguity by mapping the constraints jointly against the fingerprint space.

- *Inertial and visible light sensors.* Similarly, inertial-based systems' drifting may be corrected by visual information provided through the camera [88].

- *Visible light and electromagnetic sensors.* VL-based systems require sufficient natural light, and good camera viewing angle to function. In contrast, electromagnetic-based

TABLE 9 Comparison of notable hybrid system performance

Authors	Sensor	Fusion technique	Testbed	Accuracy	Notes
Papaioannou et al. [89]	WiFi, camera	Kalman Filter	11 × 12 m area, 1 camera mounted 10 m off the ground	1 m, 90% chance	The user trajectory is estimated through a series of WiFi scans, fused with the images provided by the ceiling mounted camera. At 60 frame window, the positioning accuracy using both WiFi and camera was three times higher than that with only the camera information.
Wang et al. [90]	Camera, magnetometer	Long-short term memory	12 × 6 m lab	3.5 m, 90% chance	Bimodal images were built to train the deep LSTM network. The magnetic field and light data were then used to estimate the user location.
Liu et al. [101]	Camera, magnetometer	Particle Filter	4,094 m^2 lab	1.34 m, 90% chance	A fingerprinting database of the magnetic field and image of each location was generated off-line. Convolutional neural network was employed to extract deep features from the images, which are then fused with the magnetic signals.
Zou et al. [102]	Inertial sensors, WiFi, Bluetooth	Particle Filter	600 m^2 office, 8 Bluetooth beacons, 8 WiFi APs	1.1 m, 90% chance	Under the same testbed, WiFi alone achieved 3.1 m, 90% chance, whereas inertial tracking quickly degraded beyond 4m after just 168 steps.
Faragher and Harle [62]	Inertial sensors, WiFi, magnetometer	Extended Kalman Filter	45 m long office corridor	2.7 m, 95% chance	WiFi and magnetic fingerprinting were used to constrain inertial sensors' drifting. The positioning error with only the inertial sensors degraded to 16 m after 520 steps.
Ban et al. [103]	Inertial sensors, WiFi, magnetometer	Gaussian mixture models	118 m walking trajectory in a station	6 m mean error, 77% floor accuracy	The WiFi and magnetic fingerprints were represented by a GMM. The positioning accuracy was slightly lower at 7.6 m mean error, 74% floor accuracy, when the phone was in pocket.
Lynen et al. [92]	Inertial sensors, camera, time-of-flight	Extended Kalman filter	1.44 km walking trajectory, 703,362 3D points	25 cm mean error	The 3D point cloud is generated off-line from a set of database images.
Gu et al. [104]	WiFi, camera, magnetometer	Weighted average	90 × 50m lab, 25 rooms	1 m, 90% chance	The users' spatio-temporal co-occurrence information were combined with the WiFi and magnetic fingerprints. Bluetooth was employed to detect users' co-location. Position information are combined for users staying in the same room.
Kim et al. [105]	WiFi, magnetometer, cellular	Weighted average	2,000 m^2 lab, 3 × 3 m^2 grid for training	3.5 m, 90% chance	A fingerprinting data of the WiFi, magnetic, and cellular signals was generated. The search space was sequentially reduced by first applying WiFi fingerprints, then cellular fingerprints, and finally the magnetic fingerprints.
Ruotsalainen et al. [106]	WiFi, Bluetooth, GNSS, inertial sensors, camera	Extended Kalman filter	Office building	3.4 m mean accuracy	The positioning accuracy in the same office, with occasional GNSS data was 13.1 m, with Bluetooth was 6.3 m, with WiFi was 8.9 m, mean error.
Nguyen and Luo [42]	Camera, time-of-flight, WiFi, magnetometer	Particle filter	900 m^2 auditorium	3.5 m, 90% chance	The positioning error in a separate office corridor was only 50 cm, 90% chance. Using only WiFi, the result was 2.5 m, 90% chance.
Du et al. [107]	Camera, magnetometer, time-of-flight	Weighted average	335 m^2 lab	1.6 m mean error	The positioning accuracy in the same office, with only the magnetic field 3.1 m, with the camera was 2.1 m, mean error.

systems use invisible waves of signal that can function well regardless of the lighting condition [89,90].

At individual sensor level within each category, some potential matchings are:

- *WiFi and magnetometer*. Distant indoor positions mostly observe different WiFi or Bluetooth fingerprints (as the wireless signals propagate and attenuate), but similar magnetic fingerprints (because of its highly spatial similarity). In contrast, neighbourhood positions tend to have similar WiFi or Bluetooth fingerprints, but different magnetic ones [42,91]. Therefore, they complement each other perfectly.
- *Camera and time-of-flight*. Camera captures a broad overview of the surroundings including colour, objects, etc. whereas ToF adds the depth dimension to the image [92].

Tables 8 and 9 summarise some notable hybrid indoor positioning systems in the recent literature.

10 | CONCLUSION

In this review, we have extensively overviewed the most relevant smartphone-based technologies for indoor positioning, and compared the performance of the most interesting positioning systems, through the reported literature and via two competitions. Throughout the journey, we have proposed novel open questions for the academic researchers, and identified some of the most practical systems with mass adoption potentials for the consumers. This final section will highlight some major points and present the roadmap for future indoor positioning systems.

10.1 | The common challenges

With more and more indoor positioning research articles being published yearly, each one improves on the predecessor's accuracy, reliability, practicality etc., we have identified the following common hurdles.

- *The trade-off between accuracy and affordability*. Most fine-grained centimetre level accuracy systems are costly, whereas affordable systems employing off-the-shelf components may only manage room-level accuracy. The ideal highly accurate and affordable system is still a work-in-progress at this stage.
- *Changing the user routine*. Most users are not willing to carry an extra piece of hardware, or having to deploy additional tracking beacons, or doing exhausting maintenance work, just to access the positioning service. However, most current systems often need one or more of these requirements.

The majority of systems in the literature often prioritise positioning accuracy as the top target for their systems.

Although this is no doubt an important criterion, there are other considerations.

- *Sensor boundary*. The hardware capability in most sensors appears to have reached its limitation (i.e. it is nearly impossible to push pure WiFi RSS-based system to consistently achieve centimetre accuracy level, in every building).
- *Algorithm efficiency*. The positioning accuracy cannot be improved endlessly with new algorithms, under the same training environment.

In short, it is challenging to rely on the same set of sensors and algorithms without any fundamental change to the problem in hand. This leads us to the '*what next?*' question for future indoor positioning researchers, to be addressed in the next section.

10.2 | The roadmap for future indoor positioning

The ultimate research question is: '*What will the future de-facto standard indoor positioning system look like?*' Such system may take some hints from the current influential GNSS, which are:

- *Ease of accessibility*. The satellite signals are ready to be accessed virtually everywhere outdoors, without any restriction.
- *Simple calculation*. The user position is computed directly by straight-forward triangulation of the satellite signals.
- *High scalability*. The system design is decentralised, so that the clients work out their position.

The futuristic equivalent indoor GNSS-like would no doubt share the above features. However, there are two fundamental differences between outdoor positioning and the indoor version. First, GNSS was designed to guide the user to the right building (e.g. shop, house) with a high tolerance of 3 m to 5 m error, whereas indoor navigation systems, at minimum, must be able to guide the user to the right room with much lower tolerance of 1 m to 2 m error. Second, GNSS assumes unobstructed line of view from the satellite to the user which is acceptable outdoors, while the complex indoor interiors make the wireless signal propagate in unexpected fashions.

As such, it may require a complete different approach or new technology for the ideal indoor positioning system. Nevertheless, technologies have evolved rapidly for the past 7 to 8 years, which fortunately has also led to the changes in the user's behaviour.

- *Smartphone as a new accessory*. People carry their smartphones everywhere outdoors, but also indoors for daily routine (e.g. email checking, Internet surfing, friend chatting). In addition, many people wear smartwatches (a small

form factor of the smartphones) while sleeping for fitness tracking. We have now embraced the smartphones as part of our inseparable necessities.

- *Smart buildings.* The indoor infrastructure are improving significantly with modern technologies. Almost every new building is well populated with WiFi APs for wireless communication, and Camera with Internet access.

The Microsoft and IPIN indoor localisation competitions, where the top academic institutions and industrial companies came together to present their systems, may offer some insights in where future indoor positioning is heading to. First, WiFi, being a ubiquitous technology in almost every modern building, looks set to stay, despite its current room-level accuracy on average. With the introduction of more advanced measurement such as RTT, the positioning accuracy challenge will be addressed in future systems. The competition results have also demonstrated that most competitors incorporate the WiFi signals in their systems. Second, smartphones have now equipped with high sampling rate IMU sensors, which can track the user's movement more accurately in a timely fashion (they were originally included for just basic app support). This observation was reflected in the two competitions where the most accurate systems all involved IMU.

In closing, the ubiquity of smartphones and improving building infrastructure in recent years have presented new opportunities to develop the ideal future indoor positioning systems, which are not only capable of fine-grained sub-metre accuracy, but are also affordable and simple to operate.

ORCID

Khuong An Nguyen https://orcid.org/0000-0001-6198-9295

REFERENCES

1. Lymberopoulos, D., Liu, J.: The microsoft indoor localization competition: experiences and lessons learnt. IEEE Signal Process. Mag. 34(5), 125–140 (2017)
2. Zafari, F., Gkelias, A., Leung, K.K.: A survey of indoor localization systems and technologies. IEEE Commun. Surv. Tutor. 21(3), 2568–2599 (2019)
3. Au, E.: The latest progress on IEEE 802.11 MC and IEEE 802.11 AI standards. IEEE Veh. Technol. Mag. 11(3), 19–21 (2016)
4. Halperin, D., et al.: Tool release: gathering 802.11 n traces with channel state information. Comput. Commun. Rev. 41(1), 53–53 (2011)
5. Chandrika, V.R.: Direction finding capability in bluetooth 5.1 standard. In: Ubiquitous Communications and Network Computing: Second EAI International Conference, Bangalore, India, February 8–10, 2019, 276, pp. 53. Springer, (2019)
6. Hofmann-Wellenhof, B., Lichtenegger, H., Collins, J.: Global positioning system: theory and practice. Springer Science & Business Media (2012)
7. Kolodziej, K.W., Hjelm, J.: Local positioning systems: LBS applications and services. CRC press CRC press, New York (2017)
8. Pu, C.C., Pu, C.H., Lee, H.J.: Indoor location tracking using received signal strength indicator. In: Emerging communications for wireless sensor networks, pp. 11. InTech, Rijeka (2011)
9. Golestani, A., et al.: Improving the precision of RSSI-based low-energy localization using path loss exponent estimation. In: 2014 11th Workshop on Positioning, Navigation and Communication (WPNC), 12-13 March 2014, pp. 1–6. IEEE, Dresden (2014)
10. Nowak, T., et al.: A path loss and fading model for RSSI-based localization in forested areas. In: 2016 IEEE-APS Topical Conference on Antennas and Propagation in Wireless Communications (APWC), 19-23 September 2016, pp. 110–113. IEEE, Cairns (2016)
11. Samimi, M.K., Rappaport, T.S., MacCartney, G.R.: Probabilistic omnidirectional path loss models for millimeter-wave outdoor communications. IEEE Wirel. Commun. Lett. 4(4), 357–360 (2015)
12. Sun, S., et al.: Synthesising omnidirectional antenna patterns, received power and path loss from directional antennas for 5g millimeter-wave communications. In: 2015 IEEE Global Communications Conference (GLOBECOM), 6-10 December 2015, pp. 1–7. IEEE, San Diego (2015)
13. Bahl, P., Padmanabhan, V.N.: Radar: an in-building RF-based user location and tracking system Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No. 00CH37064), 2, pp. 775–784. IEEE (2000)
14. Bishop, C.M.: Pattern recognition and machine learning. Springer, New York (2006)
15. Cormen, T.H., et al.: Bayesian indoor positioning systems. In: Introduction to algorithms. MIT Press (2009)
16. Madigan, D., et al.: Bayesian indoor positioning systems. In: Proceedings IEEE 24th annual joint conference of the IEEE Computer and Communications Societies, vol. 2, Miami, FL, 13–17 pp. 1217–1227. IEEE, New York March (2005)
17. Zhao, F., Huang, T., Wang, D.: A probabilistic approach for wifi fingerprint localization in severely dynamic indoor environments. IEEE Access. 7, 116348–116357 (2019)
18. Nguyen, K.A., Luo, Z.: Enhanced conformal predictors for indoor localisation based on fingerprinting method. IFIP International Conference on Artificial Intelligence Applications and Innovations, pp. 411–420. Springer, Cyprus (2013)
19. Nguyen, K.A., Luo, Z.: Reliable indoor location prediction using conformal prediction. Ann. Math. Artif. Intell. 74(1-2), 133–153 (2015)
20. Shalev-Shwartz, S., Ben-David, S.: Understanding machine learning: from theory to algorithms. Cambridge University Press, New York (2014)
21. Honkavirta, V., et al.: A comparative survey of WLAN location fingerprinting methods. In: Proceedings of the 6th Workshop on Positioning, Navigation and Communication 2009 WPNC'09, March 19, 2009, pp. 243–251. IEEE, Hannover (2009)
22. Huang, C.N., Chan, C.T.: Zigbee-based indoor location system by k-nearest neighbour algorithm with weighted rssi. Procedia Comput. Sci. 5, 58–65 (2011)
23. Shin, B., et al.: Enhanced weighted k-nearest neighbour algorithm for indoor wi-fi positioning systems. In: Computing Technology and Information Management (ICCM), 8th International Conference on, pp. 574–577. IEEE (2012)
24. Draper, N.R., Smith, H.: Applied regression analysis. John Wiley & Sons, New York (2014)
25. Bisio, I., et al.: Smart probabilistic fingerprinting for wifi-based indoor positioning with mobile devices. Pervasive Mob. Comput. 31, 107–123 (2016)
26. Murphy, K.P.: Machine learning: a probabilistic perspective. MIT Press, Cambridge (2012)
27. Youssef, M., Agrawala, A.: The Horus WLAN location determination system. In: Proceedings of the 3rd international conference on Mobile systems, applications, and services, pp. 205–218. ACM, Seattle (2005)
28. Nguyen, K.A.: A performance guaranteed indoor positioning system using conformal prediction and the wifi signal strength. J. Inform. Telecommun. 1(1), 41–65 (2017)
29. Mok, E., Retscher, G.: Location determination using wifi fingerprinting versus wifi trilateration. J. Locat. Based Serv. 1(2), 145–159 (2007)
30. Cantón Paterna, V., The Horus WLAN location determination system: A bluetooth low energy indoor positioning system with channel diversity, weighted trilateration and Kalman filtering. Sensors. 17(12), 2927 (2017)

31. Kaemarungsi, K., Krishnamurthy, P.: Analysis of WLAN'S received signal strength indication for indoor location fingerprinting. Pervasive Mob. Comput. 8(2), 292–316 (2012)

32. Machaj, J., Brida, P., Tatarova, B.: Impact of the number of access points in indoor fingerprinting localization. In: 20th International Conference Radioelektronika 2010, pp. 1–4. IEEE, Brno (2010)

33. Milioris, D., et al.: Empirical evaluation of signal-strength fingerprint positioning in wireless LANS. In: Proceedings of the 13th ACM International Conference on Modeling, Analysis, and Simulation of Wireless and Mobile Systems, 5–13 (2010)

34. Chen, Q., et al.: Placement of access points for indoor wireless coverage and fingerprint-based localization. In: 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing, pp. 2253–2257. IEEE, Zhangjiajie (2013)

35. Sharma, C., et al.: Access point placement for fingerprint-based localization. In: 2010 IEEE International Conference on Communication Systems, pp. 238–243. IEEE, Singapore (2010)

36. Nguyen, K.A.: Robot-based evaluation of bluetooth fingerprinting. Master's thesis, Computer Lab. University of Cambridge (2011)

37. Faragher, R., Harle, R.: Location fingerprinting with bluetooth low energy beacons. IEEE J. Sel. Area Commun. 33(11), 2418–2428 (2015)

38. Moghtadaiee, V., Dempster, A.G., Lim, S.: Indoor localization using FM radio signals: a fingerprinting approach. In: 2011 International Conference on Indoor Positioning and Indoor Navigation, pp. 1–7. IEEE, New York (2011)

39. Viel, A., et al.: An original approach to positioning with cellular fingerprints based on decision tree ensembles. J. Locat. Based Serv. 13(1), 25–52 (2019)

40. Nguyen, K.A., Luo, Z., Watkins, C.: On the feasibility of using two mobile phones and WLAN signal to detect co-location of two users for epidemic prediction. In: Progress in Location-Based Services 2014, pp. 63–78. Springer (2015)

41. Nguyen, K., Luo, Z.: Evaluation of bluetooth properties for indoor localisation. In: Progress in Location-Based Services, pp. 127–149. Springer (2013)

42. Nguyen, K.A., Luo, Z.: On assessing the positioning accuracy of Google tango in challenging indoor environments. In: 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8. IEEE, Sapporo (2017)

43. Li, L.: Time-of-flight camera–an introduction: Technical white paper, 1 (2014). SLOA190B

44. Lowe, D.G.: Object recognition from local scale-invariant features. Proceedings of the Seventh IEEE International Conference on Computer Vision, 2, 1150–1157. IEEE, Kerkyra (1999)

45. Bay, H., et al.: Speeded-up robust features (surf). Comput. Vis. Image Understand. 110(3), 346–359 (2008)

46. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. 60(2), 91–110 (2004)

47. Nistér, D., Naroditsky, O., Bergen, J.: Visual odometry. In: Proceedings of the 2004 IEEE computer society conference on computer vision and pattern recognition. CVPR 2004, vol. 1, pp. I–I. IEEE, Washington (2004)

48. Scaramuzza, D., Fraundorfer, F.: Visual odometry tutorial. IEEE Robot. Autom. Mag. 18(4), 80–92 (2011)

49. Yang, Z., et al.: Wearables can afford: light-weight indoor positioning with visible light. In: Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services, 317–330 (2015)

50. Li, L., et al.: Epsilon: A visible light based positioning system. In: 11th USENIX Symposium on Networked Systems Design and Implementation (NSDI '14) Seattle, WA, 2–4 April 2014

51. Fang, Z., Scherer, S.: Real-time onboard 6DoF localization of an indoor MAV in degraded visual environments using a RGB-d camera. In: 2015 IEEE international conference on robotics and automation (ICRA), pp. 5253–5259. IEEE, Seattle (2015)

52. Zhang, C., Zhang, X.: Litell: robust indoor localization using unmodified light fixtures. In: Proceedings of the 22nd Annual International Conference on Mobile Computing and Networking, pp. 230–242. ACM, New York (2016)

53. Xiao, A., et al.: An indoor positioning system based on static objects in large indoor scenes by using smartphone cameras. Sensors. 18(7), 2229 (2018)

54. Nguyen, K.A., et al.: Location tracking using smartphone accelerometer and magnetometer traces. In: Proceedings of the 14th International Conference on Availability, Reliability and Security, 1–9 (2019)

55. Nguyen, K.A., et al.: Realtime tracking of passengers on the london underground transport by matching smartphone accelerometer footprints. Sensors. 19(19), 4184 (2019)

56. Correa, A., et al.: A review of pedestrian indoor positioning systems for mass market applications. Sensors. 17(8), 1927 (2017)

57. Woodman, O.J.: An introduction to inertial navigation. University of Cambridge, Computer Laboratory, Cambridge (2007)

58. Harle, R: A survey of indoor inertial positioning systems for pedestrians. IEEE Commun. Surv. Tutor. 15(3), 1281–1293 (2013)

59. Weinberg, H.: Using the adxl202 in pedometer and personal navigation applications. Analog Devices AN-602 application note. 2(2), 1–6 (2002)

60. Solin, A., et al.: Inertial odometry on handheld smartphones. In: 2018 21st International Conference on Information Fusion (FUSION), pp. 1–5. IEEE, Cambridge (2018)

61. Kang, W., et al.: Improved heading estimation for smartphone-based indoor positioning systems. In: 2012 IEEE 23rd International Symposium on Personal, Indoor and Mobile Radio Communications-(PIMRC), pp. 2449–2453. IEEE, Sydney (2012)

62. Faragher, R., Harle, R.: Smartslam-an efficient smartphone indoor positioning system exploiting machine learning and opportunistic sensing. ION GNSS, vol. 13, pp. 1–14 (2013)

63. Nguyen, K.A., Luo, Z.: Dynamic route prediction with the magnetic field strength for indoor positioning. Int. J. Wireless Mobile Comput. 12(1), 16–35 (2017)

64. Zhao, H., et al.: Pedestrian dead reckoning using pocket-worn smartphone. IEEE Access. 7, 91063–91073 (2019)

65. Qian, J., et al.: An improved indoor localization method using smartphone inertial sensors. In: International Conference on Indoor Positioning and Indoor Navigation, pp. 1–7. IEEE, Montbeliard (2013)

66. Roy, N., Hassanieh, H., Roy Choudhury R.: Backdoor: Making microphones hear inaudible sounds. In: Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services, pp. 2–14. ACM, New York (2017)

67. Deshotels, L.: Inaudible sound as a covert channel in mobile devices. In: Proceedings of the 8th USENIX conference on Offensive Technologies, 16–16 (2014)

68. Hobbs, P.V.: Introduction to atmospheric chemistry. Cambridge University Press, Princeton (2000)

69. Li, B., Harvey, B., Gallagher, T.: Using barometers to determine the height for indoor positioning. In: International Conference on Indoor Positioning and Indoor Navigation, pp. 1–7. IEEE, Montbeliard (2013)

70. Tarzia, S.P., et al.: Indoor localization without infrastructure using the acoustic background spectrum. In: Proceedings of the 9th International Conference on Mobile systems, Applications, and Services, pp. 155–168. ACM, Maryland (2011)

71. Rossi, M., et al.: Roomsense: an indoor positioning system for smartphones using active sound probing. In: Proceedings of the 4th Augmented Human International Conference, 89–95 (2013)

72. Rishabh, I., Kimber, D., Adcock, J.: Indoor localization using controlled ambient soundsIn: 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–10. IEEE, Sydney (2012)

73. Zhang, L., et al.: Toa estimation of chirp signal in dense multipath environment for low-cost acoustic ranging. IEEE Trans. Instrum. Meas. 68(2), 355–367 (2018)

74. Cai, C., et al.: Asynchronous acoustic localization and tracking for mobile targets. IEEE Internet Things J. 7(2), 830–845 (2019)

75. Nguyen, K.A., Luo, Z.: Selective mixture of Gaussians clustering for location fingerprinting. In: Proceedings of the 11th International Conference on Mobile and Ubiquitous Systems: Computing, Networking and

Services, pp. 328–337. ICST (Institute for Computer Sciences, Social Informatics and Telecommunications Engineering), London (2014)

76. King, R.C., et al.: Application of data fusion techniques and technologies for wearable health monitoring. Med. Eng. Phys. 42, 1–12 (2017)

77. Beddar Wiesing, S., Bieshaar, M.: Multi-sensor data and knowledge fusion–a proposal for a terminology definition. arXiv preprint arXiv:200104171 (2020)

78. Wen, H., et al.: On assessing the accuracy of positioning systems in indoor environments. In: European Conference on Wireless Sensor Networks, pp. 1–17. Springer, Ghent (2013)

79. Rhudy, M.B., Salguero, R.A., Holappa, K.: A Kalman filtering tutorial for undergraduate students. Int. J. Comput. Sci. Eng. Surv. 8(1), 1–9 (2017)

80. Faragher, R.: Understanding the basis of the Kalman filter via a simple and intuitive derivation lecture notes. IEEE Signal Process. Mag. 29(5), 128–132 (2012)

81. An, D., Choi, J.H., Kim, N.H.: Prognostics 101: a tutorial for particle filter-based prognostics algorithm using Matlab. Reliab. Eng. Syst. Saf. 115, 161–169 (2013)

82. Gershman, S.J., Blei, D.M.: A tutorial on Bayesian nonparametric models. J. Math. Psychol. 56(1), 1–12 (2012)

83. Reynolds, D.A.: Gaussian mixture models. Encyclopedia of Biometrics pp. 741. MIT Press, Lexington, MA (2009)

84. Li, X., Wang, J., Liu, C.: A bluetooth/PDR integration algorithm for an indoor positioning system. Sensors. 15(10), 24862–24885 (2015)

85. Putta, R., Misra, M., Kapoor, D.: Smartphone based indoor tracking using magnetic and indoor maps. In: 2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP), pp. 1–6. IEEE (2015)

86. Zhu, N., et al.: A novel particle filter approach for indoor positioning by fusing wifi and inertial sensors. Chin. J. Aeronaut. 28(6), 1725–1734 (2015)

87. Maghdid, S.A., et al.: Indoor human tracking mechanism using integrated onboard smartphones wi-fi device and inertial sensors. Telecommun. Syst. 71(3), 447–458 (2019)

88. Al Delail B., et al.: Indoor localization and navigation using smartphones augmented reality and inertial tracking. In: 2013 IEEE 20th International Conference on Electronics, Circuits, and Systems (ICECS), pp. 929–932. IEEE, Abu Dhabi (2013)

89. Papaioannou, S., et al.: Fusion of radio and camera sensor data for accurate indoor positioning. In: 2014 IEEE 11th International Conference on Mobile Ad Hoc and Sensor Systems, pp. 109–117. IEEE, Philadelphia (2014)

90. Wang, X., Yu, Z., Mao, S.: Deepml: Deep LSTM for indoor localization with smartphone magnetic and light sensors. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, Kansas City (2018)

91. Nguyen, K.A., Watkins, C., Luo, Z.: Co-location epidemic tracking on london public transports using low power mobile magnetometer. In: 2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–8. IEEE, Sapporo (2017)

92. Lynen, S., et al.: Get out of my lab: large-scale, real-time visual-inertial localizatio. Robotics: Science and Systems, vol. XI, pp. 37 (2015)

93. Xu, Q., Zheng, R., Hranilovic, S. Idyll: Indoor localization using inertial and light sensors on smartphones. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 307–318. ACM, Osaka (2015)

94. Yang, H., et al.: Smartphone-based indoor localization system using inertial sensor and acoustic transmitter/receiver. IEEE Sensor J. 16(22), 8051–8061 (2016)

95. Shu, Y., et al.: Magicol: indoor localization using pervasive magnetic field and opportunistic wifi sensing. IEEE J. Sel. Area Commun. 33(7), 1443–1457 (2015)

96. Shao, W., et al.: Indoor positioning based on fingerprint-image and deep learning. IEEE Access. 6, 74699–74712 (2018)

97. Baniukevic, A., Jensen, C.S., Lu, H.: Hybrid indoor positioning with wi-fi and bluetooth: architecture and performance. In: 2013 IEEE 14th International Conference on Mobile Data Management, vol. 1, pp. 207–216. IEEE, Milan (2013)

98. Liu, H., et al.: Push the limit of wifi based localization for smartphones. In: Proceedings of the 18th annual international conference on Mobile computing and networking, pp. 305–316. ACM, Istanbul (2012)

99. Bisio, I., et al.: Wifi meets barometer: smartphone-based 3d indoor positioning method. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6. IEEE, Kansas City (2018)

100. Xu, H., et al.: Enhancing wifi-based localization with visual clues. In: Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing, pp. 963–974. ACM, Osaka (2015)

101. Liu, Z., et al.: Fusion of magnetic and visual sensors for indoor localization: infrastructure-free and more effective. IEEE Trans. Multimed. 19(4), 874–888 (2016)

102. Zou, H., et al.: Accurate indoor localization and tracking using mobile phone inertial sensors, wifi and ibeacon. In: 2017 IEEE International Symposium on Inertial Sensors and Systems (INERTIAL), pp. 1–4. IEEE (2017)

103. Ban, R., et al.: Indoor positioning method integrating pedestrian dead reckoning with magnetic field and wifi fingerprints. In: 2015 Eighth international conference on mobile computing and ubiquitous networking (ICMU), pp. 167–172. IEEE (2015)

104. Gu, F., Niu, J., Duan, L.: Waipo: a fusion-based collaborative indoor localization system on smartphones. IEEE/ACM Trans. Netw. 25(4), 2267–2280 (2017)

105. Kim, B., et al.: A multi-pronged approach for indoor positioning with wifi, magnetic and cellular signals. In: 2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 723–726. IEEE, Busan (2014)

106. Ruotsalainen, L., Kuusniemi, H., Chen, R.: Visual-aided two-dimensional pedestrian indoor navigation with a smartphone. J. Glob. Position. Syst. 10(1), 11–18 (2011)

107. Du, Y., Arslan, T., Juri, A.: Camera-aided region-based magnetic field indoor positioning. In: 2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–7. IEEE, Alcala de Henares (2016)