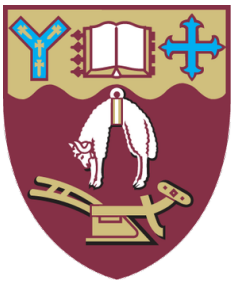
UNIVERSITY OF CANTERBURY



Visual-Inertial First Responder Localisation in Large-Scale Indoor Training Environments

Author:
Humayun Khan

Supervised by:
Prof. Robert W. Lindeman
Dr. Graeme Woodward

A thesis submitted in conformity with the requirements
for the degree of
Doctor of Philosophy
Human Interface Technology Laboratory New Zealand,
College of Engineering

July 2021

Abstract

Accurately and reliably determining the position and heading of first responders undertaking training exercises can provide valuable insights into their situational awareness and give a larger context to the decisions made. Measuring first responder movement, however, requires an accurate and portable localisation system. Training exercises often take place in large-scale indoor environments with limited power infrastructure to support localisation. Indoor positioning technologies that use radio or sound waves for localisation require an extensive network of transmitters or receivers to be installed within the environment to ensure reliable coverage. These technologies also need power sources to operate, making their use impractical for this application. Inertial sensors are infrastructure-independent, low-cost, and low-power positioning devices which are attached to the person or object being tracked, but their localisation accuracy deteriorates over long-term tracking due to intrinsic biases and sensor noise.

This thesis investigates how inertial sensor tracking can be improved by providing correction from a visual sensor that uses passive infrastructure (fiducial markers) to calculate accurate position and heading values. Even though using a visual sensor increase the accuracy of the localisation system, combining them with inertial sensors is not trivial, especially when mounted on different parts of the human body and going through different motion dynamics. Additionally, visual sensors have higher energy consumption, requiring more batteries to be carried by the first responder.

This thesis presents a novel sensor fusion approach by loosely coupling visual and inertial sensors to create a positioning system that accurately localises walking humans in large-scale indoor environments. Experimental evaluation of the devised localisation system indicates sub-metre accuracy for a $250m$ long indoor trajectory. The thesis also proposes two methods to improve the energy efficiency of the localisation system. The first is a distance-based error correction approach which uses distance estimation from the foot-mounted inertial sensor to reduce the number of corrections required from the visual sensor. Results indicate a 70% decrease in energy consumption while maintaining sub-metre localisation accuracy. The second method is a motion-type adaptive error correction approach, which uses the human walking motion type (forward, backward, or sideways) as an input to further optimise the energy efficiency of the localisation system by modulating the operation of the visual sensor. Results of this approach indicate a 25% reduction in the number of corrections required to keep sub-metre localisation accuracy. Overall, this thesis advances the state of the art by providing a sensor fusion solution for long-term sub-metre accurate localisation and methods to reduce the energy consumption, making it more practical for use in first responder training exercises.

Acknowledgements

This work would not have been possible without the support of many people who guided me along the way. I sincerely appreciate their help.

First and foremost, I would like to thank my supervisor Rob Lindeman for his continuous encouragement and support throughout my PhD. His motivation to push my limits has inspired me to become a better version of myself. Also, his ongoing support has enabled me to freely explore the research area.

I would like to thank my co-supervisor Graeme Woodward for his guidance and encouragement during difficult times. His suggestions and insightful discussions have always steered my research in the right direction.

I am thankful to Adrian Clark for his valuable insights on computer vision problems and how to approach them. I am grateful to Craig Smith for his ideas on understanding and evaluating low-power electronic circuitry. I would like to thank the examiners of this thesis for taking the time to evaluate my work.

Many thanks to the staff and students at the HIT Lab NZ and WRC, Ken, Fred, Chris, Jackie, Mel, Ryan, James, Bhuvaneswari, Rory, Zoe, Alaeddin, Kien, Yuanjie, Merel, Nikita, William, and Josh for all the experiences and good memories during my PhD.

Last but not least, I would not have been able to complete this work without the love and support from my parents and sister. They were always there to listen and cheer me up during challenging times. Thank you very much!

Deputy Vice-Chancellor's Office
Postgraduate Research Office

UNIVERSITY OF CANTERBURY
Te Whare Wānanga o Waitaha
CHRISTCHURCH NEW ZEALAND

Co-Authorship Form

This form is to accompany the submission of any thesis that contains research reported in co-authored work that has been published, accepted for publication, or submitted for publication. A copy of this form should be included for each co-authored work that is included in the thesis. Completed forms should be included at the front (after the thesis abstract) of each copy of the thesis submitted for examination and library deposit.

Please indicate the chapter/section/pages of this thesis that are extracted from co-authored work and provide details of the publication or submission from the extract comes:

Chapter 4 *of the thesis is a reproduction of the work undertaken and published in collaboration with Adrian Clark, Graeme Woodward, and Robert W. Lindeman. Results of this work have been presented at the following journal.*

Khan, Humayun, Adrian Clark, Graeme Woodward, and Robert W. Lindeman. (2020). "Improved position accuracy of foot-mounted inertial sensor by discrete corrections from vision-based fiducial marker tracking." Sensors 20, no. 18, 5031.

Please detail the nature and extent (%) of contribution by the candidate:

The localisation system described in the paper was developed under the supervision of Robert W. Lindeman, Graeme Woodward, and Adrian Clark. Robert W. Lindeman also helped me to conduct the experiment. All the authors reviewed and edited the paper. My co-authors total contribution towards the paper did not exceed 10% of the final paper.

Certification by Co-authors:
If there is more than one co-author then a single co-author can sign on behalf of all
The undersigned certifies that:

- The above statement correctly reflects the nature and extent of the Doctoral candidate's contribution to this co-authored work
- In cases where the candidate was the lead author of the co-authored work he or she wrote the text

Name: *Robert W. Lindeman* Signature:  Date: 2021-07-26

Contents

List of Figures

List of Tables

Abbreviations

AoA	**A**ngle **o**f **A**rrival
APs	**A**ccess **P**oints
GNSS	**G**lobal **N**avigation **S**atellite **S**ystem
IMU	**I**nertial **M**easurement **U**nit
INS	**I**nertial **N**avigation **S**ystem
KF	**K**alman **F**ilter
MEMS	**M**icro **E**lectro **M**echanical **S**ystem
MM	**M**ap **M**atching
PDR	**P**edestrian **D**ead **R**eckoning
PF	**P**article **F**ilter
RFID	**R**adio **F**requency **I**dentification
RPs	**R**eference **P**oints
RSS	**R**eceived **S**ignal **S**trength
SHS	**S**tep and **H**eading **S**ystem
TDoA	**T**ime **D**ifference **o**f **A**rrival
ToA	**T**ime **o**f **A**rrival
ToF	**T**ime **o**f **F**light
UWB	**U**ltra-**W**ide**b**and
VLC	**V**isible **L**ight **C**ommunication
WLAN	**W**ireless **L**ocal **A**rea **N**etwork
ZUPT	**Z**ero **V**elocity **U**pdate

Chapter 1

Introduction

1.1 Motivation

During first responder training exercises, localisation and tracking of participants can be a useful tool for evaluating performance, coordinating movements, and improving safety and efficiency. These exercises often take place in large-scale indoor environments with limited power infrastructure, and the participants involved often have to wear protective (e.g., fireproof) suits and carry heavy equipment. Figures 1.1, 1.2, and 1.3 show a typical training environment for firefighters, military personnel, and policemen respectively. Past research has indicated that for localisation information to be useful for first responders, it has to be within a sub-metre accuracy range [1–4]. Considering these limitations, technologies for localisation should be carefully chosen to fulfil the accuracy requirements and minimise energy consumption, the infrastructure required, size, weight, and cost.

FIGURE 1.1: Firefighter urban training environment

FIGURE 1.2: Police training environment [5]

FIGURE 1.3: Military personnel training environment [6]

Many existing localisation systems are based on radio and sound waves, such as ultra-wideband [7], radio frequency identification [8], bluetooth beacons [9], global navigation satellite systems (Global Positioning System (GPS), GLObal NAvigation Satellite System (GLONASS), BeiDou, and Galileo [10]), Wi-Fi [3], and ultrasound [11]. However, radio and sound signals get attenuated by multipaths and reflections in indoor environments. To counter this and to achieve higher localisation accuracy many radio/sound transmitters and receivers have to be installed, which increases the infrastructure, cost, and power requirements. This limits the use of radio and sound waves for the training

exercise application. A more suitable tracking technology is micro-electro-mechanical system (MEMS)-based inertial sensors, as they are infrastructure independent, low-cost, and consume less power. However, inertial sensors accumulate drift (error) rapidly, which quickly decreases the localisation accuracy beyond the required sub-metre limit. Many researchers in the past have found that the rapid error growth can be checked by mounting the inertial sensor on the foot and applying a zero velocity update (ZUPT) method [12–15]. Figure 1.4 illustrates when ZUPT is applied during the stance and swing phase of walking to minimise growth in the localisation error. The ZUPT algorithm slows the rate of error growth, but the localisation accuracy still goes beyond one metre in the longer term, as inertial sensors are inherently affected by biases and sensor noise.

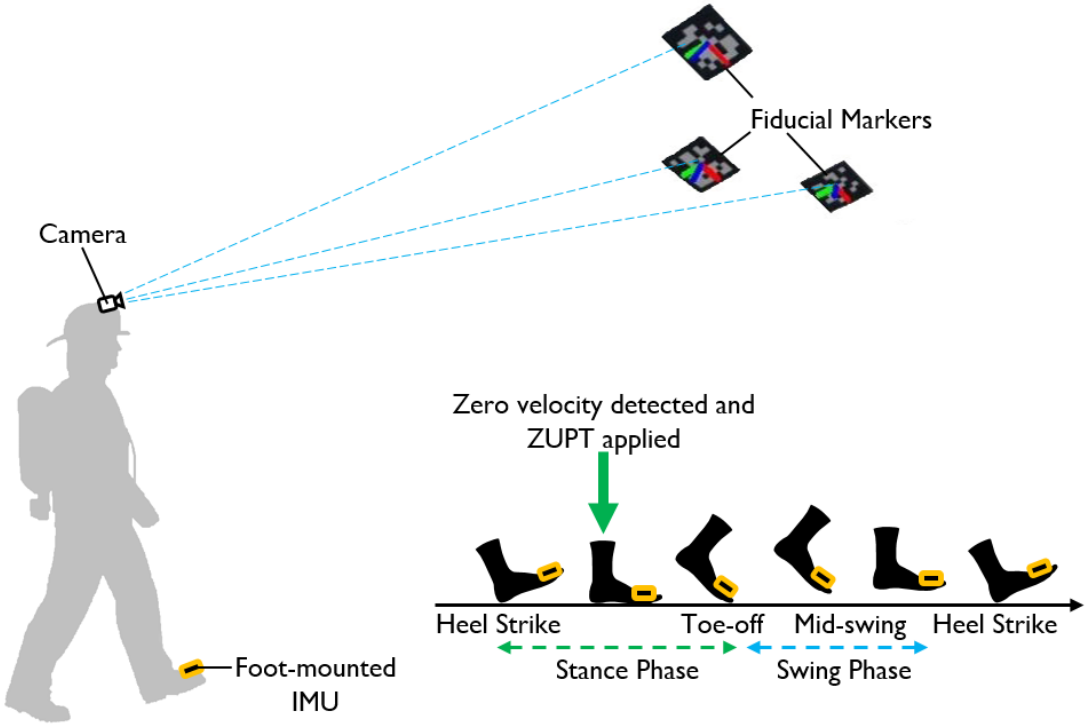


FIGURE 1.4: Localisation of first responders using a foot-mounted IMU and visual sensor.

To further check the error growth, another localisation technology that fulfils the sub-metre accuracy requirement needs to be fused with inertial sensors. With recent developments in computer vision research, camera-based localisation systems can estimate localisation information to a centimetre level of accuracy depending on the detectable visual features present in the environment [16–19]. The cameras are also low-cost devices,

and small and light enough to be carried by first responders. However, indoor environments, especially training environments, are composed of either textureless surfaces or repetitive patterns which lack unique visual features and can thus drastically decrease the localisation accuracy [20–22]. This can be resolved by installing fiducial markers in the environment, as these are passive infrastructure and do not require any energy source to operate. Figure 1.4 illustrates how a visual sensor uses fiducial markers to compute the localisation information. The terms "camera" and "visual sensor" are used interchangeably in this thesis. Training environments can be controlled and organised beforehand, which presents an opportunity to augment the environment with markers and accurately map it.

Apart from the dependence on visual features, a camera-based localisation system also has higher energy consumption than inertial sensors. A typical low-power camera consumes 1.0 W to 1.2 W [23, 24], compared to 40 to 60 mW by inertial sensors [25–29]. The difference between the two is 16 to 30 times. For portability, the localisation system for first responders is powered by electrical batteries. If the system uses a lot of energy, more batteries need to be added, or it has to be recharged more frequently. Additional batteries would increase the weight to be carried by first responders, which is not ideal as they are already wearing a heavy suit and carrying heavy equipment. Further increasing the weight could impede them from performing the exercise tasks. The other option is recharging more frequently, which is also not viable as the training exercise has to be stopped to recharge the batteries. The effect of the camera's high power consumption can be managed by intelligently controlling the camera operation. Besides **long-term localisation accuracy**, the research focus of this thesis is also to develop and evaluate methods to **decrease the overall energy consumption** of the localisation system for long-term operation.

Sensor fusion between visual and inertial sensors can ensure a cost-effective and long-term positioning solution with reduced power consumption. However, both sensors perceive changes in the position and heading values differently, as each has fundamentally different sensing characteristics (such as sampling rate, data communication, and susceptibility to noise). This makes the sensor fusion of the two sensors a far from trivial problem, especially when the sensors are mounted on two different parts of the human body. While a lot of research has been undertaken on rigidly attached visual-inertial

sensor fusion systems[30–32], much less has been explored on non-rigidly attached visual-inertial sensor fusion systems. This is challenging as each sensor goes through different motion dynamics. To address this, we devised a novel sensor fusion algorithm by loosely coupling the visual and inertial sensors with an extended Kalman filter. Another under-researched area is the energy efficiency of visual-inertial localisation systems. This is a critical issue for training exercises, as the localisation system used by first responders has a limited battery supply. There are two methods proposed in this thesis to improve the energy efficiency of the localisation system. The first method is a distance-based error correction approach, which uses distance estimation from the inertial sensor to reduce the number of corrections required from the visual sensor. The second method is a motion-type adaptive error correction approach, which uses the human walking motion types (forward, backward, and sideways) as an input to modulate the correction from the visual sensor. The overall focus of this thesis is a low-cost and sub-metre accurate localisation system that can track first responders for a longer duration of time (2 to 3 days) in large scale indoor environments for an after-action review application (non-real-time), with a limited battery supply to be carried by the first responders, and no powered infrastructure (e.g., active beacons) installed in the environment.

1.2 Applications

Our system primarily targets first responder training exercises. However, it could also be extended to the following applications:

- **Visually impaired navigation:** People with visual impairment often want to know their current location in an unfamiliar building. They have to either ask other people in the building or bring a guide along. Buildings could introduce symbols or fiducial markers, similar to Braille signage, that could be recognised by the localisation system worn by the blind person. This would help them navigate indoor environments more easily and prevent the feeling of being lost.

- **Guided tours:** Museums often lack an interactive storytelling experience. Each artefact could have a fiducial marker on the description plate for navigation and a guided tour.

- **Trade shows:** It is often hard to locate booths in large trade shows. Having fiducial markers or QR codes on each stall and a localisation application would make it easier for visitors to find the booth they want to visit.

- **Large grocery stores:** Due to the sheer size of some grocery stores, sometimes it is hard to find an item on a shopping list. Stores could have a navigation application and fiducial markers installed on aisles and different sections to guide the visitors. This would improve the shopping experience and make it easier for visitors to locate items.

- **Large shopping malls:** In large shopping malls, it is easy to lose direction. Instead of looking around for a floor map, a fiducial marker or a QR code installed on the entry of the shops and a localisation application could help a shopper locate the current position and direction to the intended shop.

1.3 Research Objectives

As described earlier in Section 1.1, the main aim of this thesis is to develop a sub-metre accurate, low-cost, and low-power localisation system for first responders which can work reliably in large-scale indoor environments with limited infrastructure and for an extended period. To achieve the main goal of our research, several minor objectives have been defined as follows:

- The first objective is to determine the mechanics of the loosely coupled sensor fusion method which fuses the data from the visual and the inertial sensors mounted on different parts of the human body. Both sensors experience the change in position and heading of the person; however, each sensor has its own perspective.

- The second objective is to investigate how the reduction in usage of visual sensors affects the accuracy and energy efficiency of the localisation system.

- The third objective is to study how the variation in human walking motion types affects the localisation accuracy, and also, how these variations can be used to further optimise the energy efficiency of the localisation system.

1.4 Publications

The research that has been presented in this thesis has resulted in the following publications. A journal publication on the loosely coupled sensor fusion approach (Chapter 4) has been published in [33]:

- Humayun Khan, Adrian Clark, Graeme Woodward, and Robert W. Lindeman. "Improved position accuracy of foot-mounted inertial sensor by discrete corrections from vision-based fiducial marker tracking." Sensors 20, no. 18 (2020): 5031.

Another journal publication will be prepared using the research presented in Chapter 5 (distance-based error correction method and its effect on the localisation accuracy and the energy efficiency) and Chapter 6 (motion type adaptive approach and the impact on the localisation accuracy and the energy efficiency).

The author has also been a co-author on publications that are not included in this thesis but are in related fields:

- Rory MS Clifford, Humayun Khan, Simon Hoermann, Mark Billinghurst, and Robert W. Lindeman. "Development of a multi-sensory virtual reality training simulator for airborne firefighters supervising aerial wildfire suppression." In 2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good), pp. 1-5. IEEE, 2018. [34]
- Rory MS Clifford, Humayun Khan, Simon Hoermann, Mark Billinghurst, and Robert W. Lindeman. "The effect of immersive displays on situation awareness in virtual environments for aerial firefighting air attack supervisor training." In 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR), pp. 1-2. IEEE, 2018. [35]
- Daniel Barry, Munir Shah, Merel Keijsers, Humayun Khan, and Banon Hopman. "XYOLO: A model for real-time object detection in humanoid soccer on low-end hardware." In 2019 International Conference on Image and Vision Computing New Zealand (IVCNZ), pp. 1-6. IEEE, 2019. [36]

1.5 Structure of the Thesis

The remainder of this thesis is structured as follows:

Chapter 2 surveys existing positioning technologies and discusses the current state of the art research on indoor localisation for pedestrians. This chapter also introduces the sensor fusion approaches and the choice of an extended Kalman filter for our system.

Chapter 3 presents the fundamentals of inertial and camera-based localisation systems, including the different coordinate systems, types of error present in inertial sensors and methods to reduce these errors, the mathematical foundations of camera-based localisation system, and the method used in this thesis to calibrate the camera.

Chapter 4 describes the development and evaluation of the loosely coupled sensor fusion approach. Localisation from the foot-mounted inertial sensors and the camera-based localisation system for the developed system is also explained in this chapter. The chapter also gives detail on accurate fiducial marker mapping in large-scale environments, ground truth methods, performance metrics, and the hardware used for this thesis.

Chapter 5 explains the distance-based error correction method that decreases the overall energy consumption of the system. Experiments and instruments used to quantify the energy consumption of the system are also described in this chapter.

Chapter 6 describes the formulation and evaluation of the motion type adaptive approach which changes the error correction mechanism with a change in motion type.

Chapter 7 summarises the thesis, identifies the contributions, and discusses possible future research directions.

Chapter 2

Indoor Positioning Technologies and Sensor Fusion Approaches

This chapter surveys the existing indoor positioning technologies utilised for localising users in an indoor space. To provide an understanding of each technology, the chapter briefly describes the methods used to calculate position and orientation from sensor measurements (Section 2.1). A comparative analysis of the technologies is also provided to examine the strengths and weaknesses of each localisation technology, and their potential application to the developed system for this thesis (Section 2.2). The localisation solution can be improved by merging technologies with complementary characteristics or two different strengths. Therefore, Section 2.3 discusses the sensor fusion methods currently being used to combine different positioning technologies, and the rationale of using an extended Kalman filter for our application is explained.

2.1 Positioning Technologies

There have been several indoor positioning systems developed in the last four decades using different technologies and types of sensors. In this section, the main indoor positioning technologies which have been used to localise humans in large-scale indoor environments will be discussed [37]. A list of these technologies is as follows:

- Camera-based Localisation
- Inertial Sensors

- Wireless Local Area Network
- Radio-Frequency Identification
- Ultra-wideband
- Bluetooth Beacons
- Ultrasound
- Visible Light Communication
- Digital Differential Barometer
- Digital Compass

2.1.1 Camera-based Localisation

Camera-based localisation for first responders can be achieved by both outside-in or inside-out tracking [38]. Outside-in tracking involves having multiple cameras installed on the infrastructure and detecting human movement by combining information from all the installed cameras. To cover larger areas such as buildings, outside-in tracking requires an installation of large number of cameras, which also need a power source. This requirement makes it less suitable for our intended application. In contrast, with inside-out tracking, one or more cameras are mounted on the user and estimate the camera pose (position and orientation) by detecting feature points or image matches in the environment. Feature points occur naturally, and are extracted through image processing [16, 17, 19], or they can be added into the environment in the form of fiducial markers [39]. Image matches are a database of images of the environment that have been stored with corresponding location information. Algorithms known as Simultaneous Localisation and Mapping (SLAM) simultaneously add feature points or image matches into a map and find the localisation information of the camera. SLAM algorithms can use input from other sensors as well, such as Lidar, infrared-based time of flight (ToF) cameras, or an additional camera to make a stereo camera system; however, these additional sensors consume significantly more power and increase the cost of the system. Therefore, we will only consider single camera methods for localisation. SLAM algorithms which uses feature points to localise and create maps are categorised as indirect SLAM, and the ones based on image matches are known as direct SLAM [40, 41].

Direct SLAM still uses the structure from motion to extract 3D geometry of the scene. Similar to indirect SLAM algorithms, direct SLAM have keyframes storing the image along with camera pose.

SLAM approaches can obtain high positioning accuracy; however, they require a lot of computational, energy, and memory resources [18, 42]. When the localisation area becomes larger, SLAM methods need to store more feature points and keyframes, which increases the memory space required. Also, during relocalisation the new query image has to be compared with all stored feature points or keyframes, which requires a lot of computational and memory resources, in turn resulting in more energy consumption. The new machine learning methods, such as convolution neural network (CNN)-based DSAC [43] and PoseNet [44], has decreased the number of feature points needed for accurate relocalisation, but they still fail in indoor environments as found by Taira et al. [20]. The indoor spaces are harder to localise than the outdoors', as large parts (walls and ceilings) are textureless and even the textured areas are often concentrated in small regions which results in unstable pose estimates [20]. Furthermore, indoor environments, especially training environments, are often made up of repetitive elements such as similar corridors, rooms, doors, and furniture which could result in incorrect feature or image matches, and in turn lead to erroneous position and heading estimates. Due to these limitations, SLAM approaches are not suitable for localising first responders with limited battery supply.

Localisation in challenging indoor environments which lack feature points can be addressed by installing fiducial markers. Fiducial markers contain feature points and are designed to be easily detectable and processed, as less image processing is required. Calculating the camera pose from fiducial markers is explained in Section 4.2.2. Examples of some of the commonly used fiducial markers are shown in Figure 2.1. With fiducial markers, the relocalisation problem also becomes efficient, as each fiducial marker has a unique pattern and identification that provides a pre-coded absolute position and heading reference. There are also some disadvantages of using fiducial markers. The markers need to be installed in the environment, so they are only appropriate for applications that can be pre-mapped, such as the controlled training environments. Another disadvantage is the visual appearance of the markers, which could be distracting for applications where aesthetics are important, such as shopping malls or hospitals. However,

for training environments, the visual appearance is of less concern than the accuracy of the localisation system.

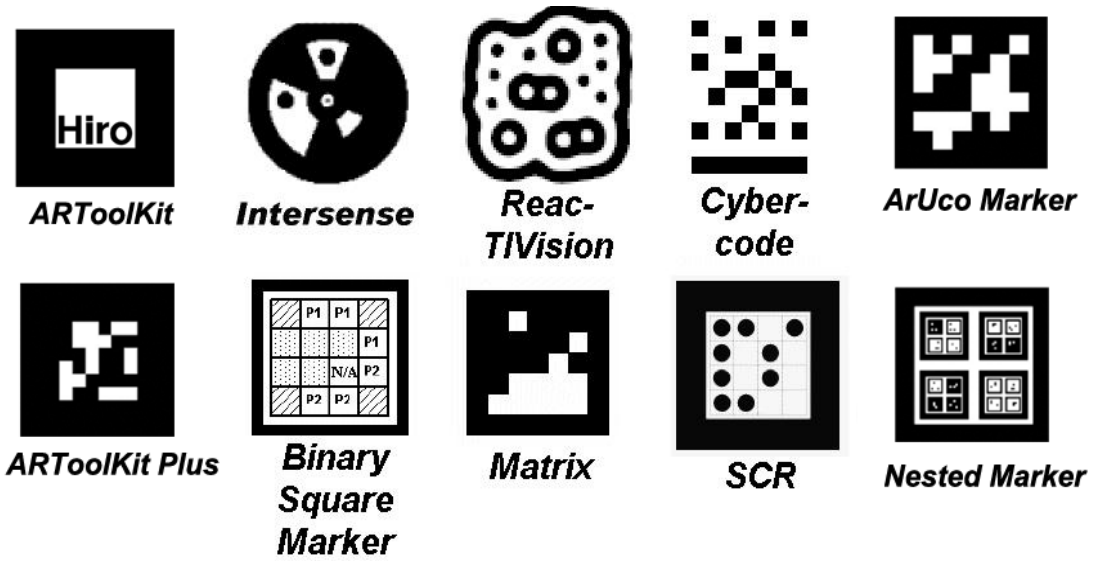


FIGURE 2.1: Commonly used fiducial markers for camera pose tracking[45]

In past research, localisation systems based on fiducial markers have been able to achieve *cm* to *dm* level of accuracy. Goronzy et al. used ceiling-mounted fiducial markers at a distance of 2 m, and they were able to achieve positioning accuracy of 2.5 *cm* [46]. Another system developed by Lee et al. with ceiling fiducial markers placed 2.2 m away was able to achieve near 6 *cm* of accuracy [47]. Marques et al., system used a Google tango phone in a 34 m^2 room, when recognising markers placed on a table [48]. They reported a 6.7 *cm* positioning error. Results from this past work strongly indicate the use of fiducial markers for an accurate positioning system.

2.1.2 Inertial Sensors

Inertial sensors measure the linear acceleration and rotational velocity of the object that they are attached to. They can track any moving object's position and orientation relative to a known starting point. There are two types of inertial sensors: accelerometers and gyroscopes. An accelerometer senses the acceleration of the object, while a gyroscope senses the angular velocity. An accelerometer and gyroscope are usually combined together on a single platform to create a device known as an inertial measurement unit (IMU). An IMU typically consists of a triad of accelerometers and a triad of gyroscopes,

mounted along three orthogonal axes to capture six degrees of freedom (DOF) motion. There are also two categories of IMUs: stable platform IMUs, and strapdown IMUs [49].

IMUs with stable platforms are more accurate, as the inertial sensors are mounted on a stable platform which is actively aligned with the navigational frame of reference (detailed in Section 3.1). Figure 2.2 illustrates the stable platform IMU. The active alignment simplifies the position and orientation calculation. The orientation of the IMU is given by reading adjacent gimbals and the position by double integrating the accelerometer signal without subtracting the gravity vector. The simple computation also leads to more accurate position and orientation values. However, the stable platform systems are bulky and expensive, and are not suitable for applications requiring mobility [50].

Strapdown IMUs on the other hand do not require gimbals to actively align the axis and can be simply strapped to the tracked entity (robot or human). Figure 2.2 illustrates a strapdown IMU. Unlike stable platform IMUs, its methods to calculate position and orientation are complex. For the position, acceleration has to be transformed to a navigational frame of reference, then a gravity vector has to be subtracted before double integration. Similarly, angular velocity has to be transformed to a navigational frame of reference before orientation is computed. Global orientation has to be constantly tracked to compute the transformation from body to a navigational frame of reference. The process is explained in more detail in Section 4.1. Due to the complexity in calculation, strapdown IMUs are less accurate; however, they are smaller, lighter, and cheaper, which makes them suitable for low-cost applications requiring mobility.

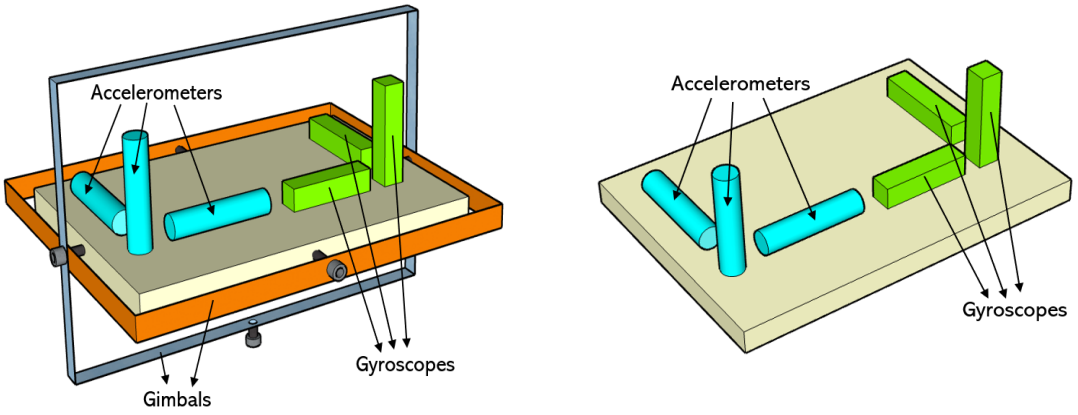


FIGURE 2.2: A stable platform IMU on the left, and a strapdown IMU on the right

Another aspect of the inertial sensors is that they are prone to error, which could be systematic (deterministic) or random (stochastic) in origin. Systematic errors include biases, scale factor, input-output non-linearity, and cross-coupling of sensitive axes measurements; while random errors in the inertial sensor data are white sensor noise, run-to-run bias, scale factor instability, and bias drift. Systematic errors can be modelled mathematically and compensated through calibration. Conversely, random errors are hard to estimate and can only be stochastically approximated. More details on these errors along with the calibration techniques are in Section 3.2.1.

Inertial sensors can also localise humans by using algorithms known as pedestrian dead reckoning (PDR). PDR can be implemented in two ways; an inertial navigation system (INS)-based approach, and a step and heading system (SHS)-based approach [51]. The difference between the two is demonstrated in Figure 2.3. The INS-based approach tracks the three dimensional position and orientation of the inertial sensor by double integrating the acceleration and angular velocity measured on each axis of the IMU. The SHS-based approach tracks the two dimensional position and orientation of the inertial sensor by accumulating step length and heading [52–54]. There are also some similarities between the two PDR methods. Both PDR estimate position and heading values relative to a known starting point. Also, the orientation is estimated by integrating the angular velocity of the IMU. However, for the INS-based PDR the orientation is in three dimensions and for the SHS-based approach in two dimensions. The INS-based and SHS-based PDR methods are described further in the next part of this section.

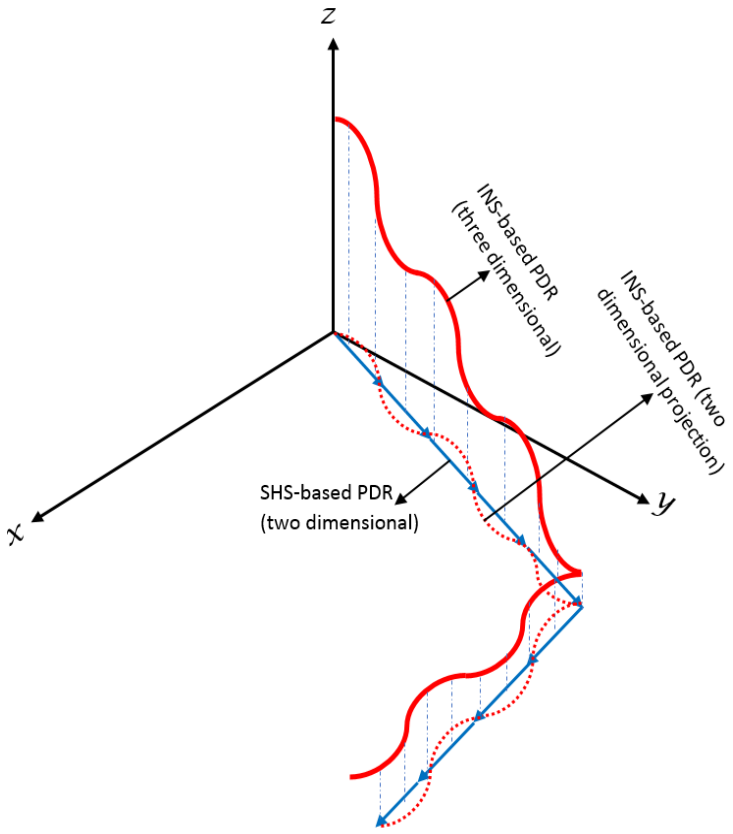


FIGURE 2.3: The INS-based PDR is indicated by the solid red line and its projection on the xy-plane by the dotted red line. The SHS-based PDR is shown by the blue line [51]

An **INS-based PDR** computes the position value by double integrating the linear acceleration calculated from the accelerometer data, and orientation by integrating the angular velocity derived from the gyroscope data. However, this approach is only accurate for short intervals and accumulates a large drift error over a longer period. To improve the overall tracking accuracy the error has to be minimised, which is achieved by applying external constraints on the system. The most widely used constraint for tracking a moving pedestrian is the zero velocity update (ZUPT) [12, 15]. This can be applied when the IMU is detected to be in a stationary state, which happens during the stance phase (shown in Figure 1.4) of the human foot motion. For the stationary state to be detected accurately, the IMU has to be mounted on the foot [15, 55]. The ZUPT algorithm performs a zero velocity update by subtracting the estimated errors in velocity and position values from the current INS estimate. Figure 2.4 illustrates the INS-based PDR algorithm with the ZUPT applied. The ZUPT algorithm is regarded as the most robust and accurate method to check the IMU drift error, and is also applicable to different motion types, such as walking (forward, sideways, backward), running,

crawling, and climbing [14, 56, 57]. Considering its accuracy and applicability to a wider range of motion, it was the chosen PDR approach for our localisation system.

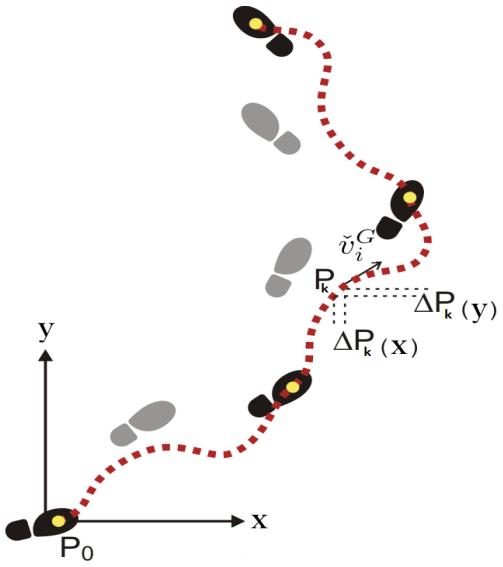


FIGURE 2.4: INS-based PDR Algorithm: P_0 is the starting point, and P_k is the position at time k which is $\Delta P_k(x, y)$ from the last point [58]

The **SHS-based PDR** calculates step lengths indirectly from the accelerometer or/and gyroscope data and step heading from the gyroscope data. It accumulates step lengths to estimate the user position. The SHS method is illustrated in Figure 2.5. The first step in the SHS-based approach is to detect steps in the data, then estimate the length of these steps (SL_1, SL_2, SL_3 in Figure 2.5), and finally, estimate the heading direction of the step ($\theta_1, \theta_2, \theta_3$ in Figure 2.5). There are several algorithms to detect steps using IMU data. These are mainly based on data processing techniques: filtering, magnitudes, and thresholding. The devices for step detection in most prior research are accelerometers [59–62] or gyroscopes [63]. There is also research on using non-inertial sensors to detect steps, such as pressure sensors [64], impact switches [65], and electromyography (EMG) sensors [66]. However, these solutions require additional hardware which increases the complexity and power consumption of the system, reduces portability and provides little advantage compared to the inertial sensor only solutions. The are two major disadvantages of the SHS-based PDR approach; first, it assumes forward motion, which means each detected step will be added in a forward direction even if it was taken in sideways or backward directions. It is still able to detect the change in direction when the whole body turns using the gyroscope data, but if the body orientation hasn't changed, such as sideways walk or backward walk, then the SHS-based PDR algorithm fails. Second, it

can only be applied to walking motion. These two limitations restrict the applicability of the SHS-based approach for our system. Therefore, we chose the INS-based PDR method for our localisation system.

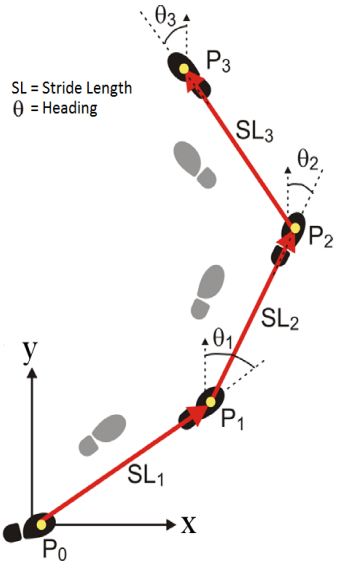


FIGURE 2.5: The SHS-based PDR Algorithm: P_0 is the starting point, and the next points (P_1, P_2, P_3) are calculated with step lengths (SL_1, SL_2, SL_3) and step heading direction ($\theta_1, \theta_2, \theta_3$) [58]

2.1.3 Wireless Local Area Network

WLAN infrastructure is installed in indoor environments to provide wireless internet connectivity to the users. The signal received from WLAN routers/access points can also be used to calculate the position value. There are five main methods that can be used for positioning from WLAN or other radio signals [67]:

- **The Received Signal Strength Indicator (RSSI)** method uses a signal propagation model to infer the position of the receiver in the environment. The signal strength value decreases with the distance from the transmitter. The distance calculated from the RSSI can also be used for trilateration when there is more than one transmitter or access point.
- **Fingerprinting** has two phases: offline and online. Both phases are illustrated in Figure 2.6. The offline phase involves surveying the indoor environment and recording the signal strength values at different locations in a database. The online phase is the localisation phase when the newly measured signal strength value is compared with the database to estimate the position.

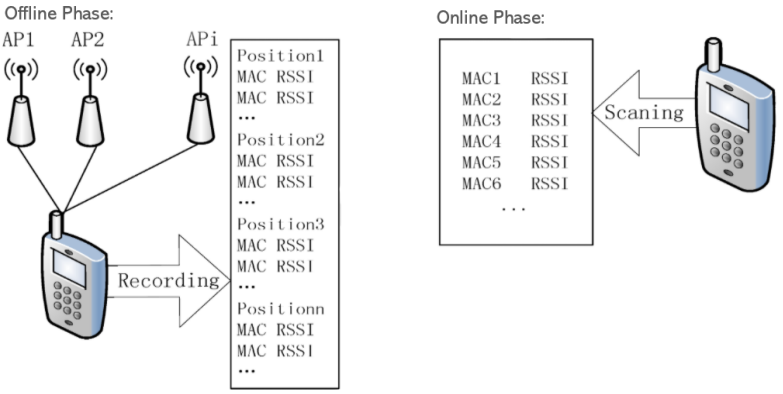


FIGURE 2.6: Fingerprinting method, the offline and the online phase [68]

- **The Cell of Origin (CoO)** divides the positioning area into cells and locates the position of the receiver from the detected transmitter or access point. The centre of each cell is a transmitter.

- **Triangulation** (Figure 2.7b) uses the angle of arrival (AoA) (Figure 2.7a) of incoming signals from three different transmitters or access points with known position values to calculate the position of the object. The AoA measurement and Triangulation are illustrated in Figure 2.7.

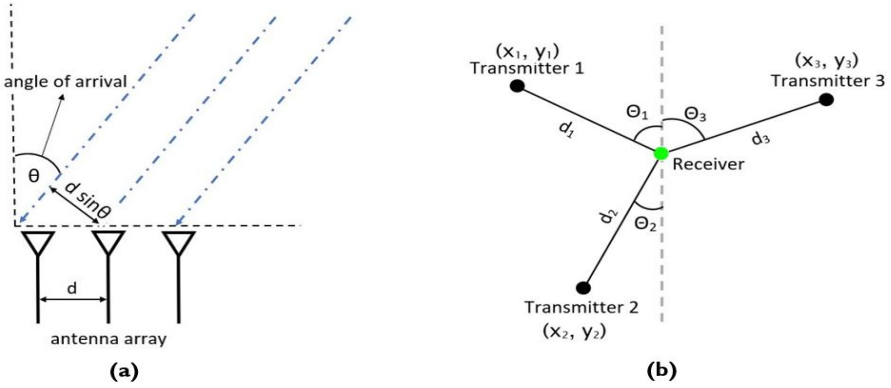


FIGURE 2.7: (a) AoA, (b) Triangulation

- **Trilateration** is based on the measured distance between the object and three or more access points. Trilateration can be done with three different methods: time of arrival (ToA), which is also known as time of flight (ToF), time difference of arrival (TDoA), and received signal strength indicator (RSSI), which was explained earlier. Trilateration with ToA and RSSI are shown in Figure 2.8a, and trilateration with TDoA is shown in Figure 2.8b. The ToA/ToF method calculates distance by measuring the time taken by the signal to propagate from the transmitter to

the receiver and the known signal velocity. For example, radio frequency signals propagate at the speed of light ($\approx$ 300,000 km/s), and take approximately 17 ns to travel 5 m. By comparison, a sound wave travels at 0.343 km/s (at a temperature of 20 $°C$), and takes about 15 ms to cover the $5m$. As a result, the radio frequency-based ToA method requires a very high resolution clock to synchronise the transmitter and the receiver compared to sound waves. Unlike ToA, the TDoA method does not require clock synchronisation between the transmitter and the receiver; however, two or more transmitters are needed to calculate the distance to the receiver. The transmitters send the signal simultaneously and the distance is calculated using the time difference of the received signals. As the signal has to be transmitted simultaneously, TDoA requires transmitters to be synchronised, which again involves highly accurate clocks.

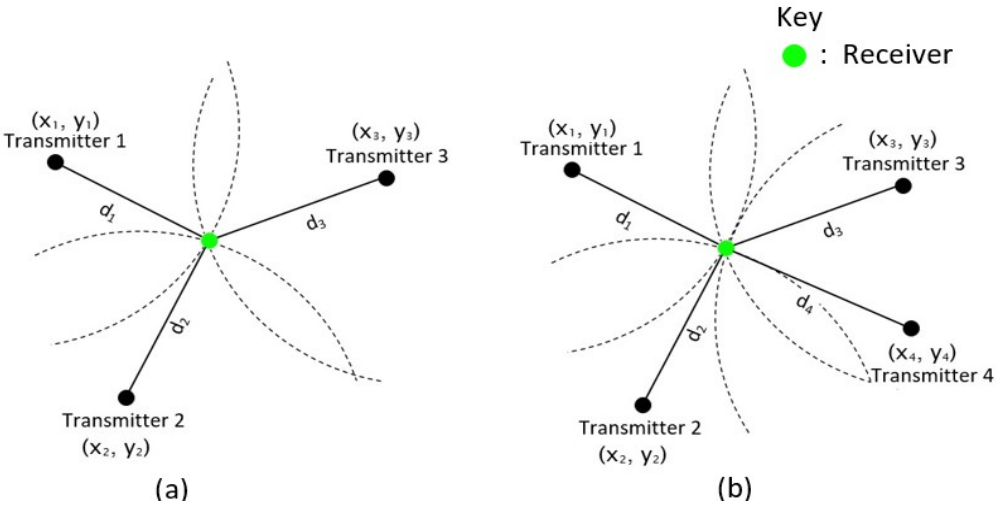


FIGURE 2.8: (a) Trilateration using the ToA and RSSI methods, (b) Trilateration using the TDoA method

From the methods described above only two have been commonly used in the literature for WLAN-based positioning: RSSI and fingerprinting. Fingerprinting is the most accurate of the two [69]; however, recording signal strength (finger print) values is a time intensive manual process. Additionally, the signal strength values are easily affected by changes in the environment, such as a movement of any object in the environment [70] or a presence of humans, which can reduce the signal strength by 10 dB [71]. This makes the WLAN less robust, and ineffective to localise human users. Hence, it was not considered for our application.

2.1.4 Radio-Frequency Identification

Radio-Frequency Identification (RFID) localisation provides absolute position information, by using a passive or active RFID tag and a reader. The tag has a unique identification which carries the position information, and the RFID can be active (equipped with an internal battery) or passive (powered by electric induction from the reader). Active tags can be read up to a range of 70 m [72] or 100 m [73], whereas passive tags are limited to a reading range of a few cms, or with a more power consuming reader up to 2 m [74, 75]. The RFID localisation system can also be fused with PDR to provide regular fixes along the pedestrian trajectory.

For passive RFID tags, the reader has to be in close proximity to the tag to detect the tag ID and location information. However, for active RFID tags the location is estimated by the received signal strength (RSS) method. The RSS method gives distance between the reader and tag using the path loss model [76], which is described by:

$$RSS(d) = RSS_0 - 10\alpha \log_{10} * \frac{d}{d_0} + \mu, \tag{2.1}$$

where RSS is measured in dBm from the RFID tag located at distance d from the reader. RSS_0 in Equation 2.1 is a received signal strength value at reference distance d_0, and is determined during calibration. α is the path loss exponent, and it primarily captures the multipath interference and obstruction in an indoor environment. μ is the Gaussian distributed noise with variance σ^2, and has a fixed value. For RSS, the system has to be calibrated once during installation to find RSS_0, α, and σ^2. The calibration does not need to be repeated unless the system is moved to another location.

In past research, various RFID-based systems have been studied for localisation. Koutsou et al. created an application with 21 active RFID tags placed in an area of 250 m^2 made up of three rooms [77]. RSS was used for positioning, and a mean error of 3.25 m was reported. In other work, Seco et al. placed 71 active RFID tags in an area of 1600 m^2 made up of 55 rooms. Using RSS, they were able to achieve a median accuracy of 1.5 m [78]. RFID is another technology besides vision that offers passive infrastructure installation; however, the detection range is limited to only a few cm which makes it harder to reliably position mobile agents. The range could be increased by using more

power consuming readers, but this also leads to higher energy consumption, which reduces battery duration. Therefore, RFID technology is not appropriate for applications requiring longer term positioning.

2.1.5 Ultra-wideband

Ultra-wideband (UWB) localisation uses short pulses (sub-nanosecond scale) of impulse radio signals to calculate UWB tag position in an indoor environment. There are four methods used to compute the position from UWB signals: time difference of arrival (TDoA), time of arrival (ToA), angle of arrival (AoA), and received signal strength indicator(RSSI), which have been explained in Section 2.1.3. Due to its accuracy (20 cm in line of sight [79]) and robustness, UWB has been implemented in several commercial indoor positioning solutions including Ubisense, DecaWave, and BeSpoon [80]. However, the cost of a UWB positioning system still remains very high, as precision clocks are required at both the transmitter and the receiver end to accurately measure the time taken by the transmitted signal. Also, UWB requires a power source at both receiver and transmitter end, which necessitates the installation of battery or electrical power infrastructure at the site. The cost and to some extent the active infrastructure requirement prevent them from being usable for applications requiring low-cost, low-power, and large-scale indoor positioning.

2.1.6 Bluetooth Beacons

Bluetooth beacons calculate position with the same 2.4 GHz radio frequency as WLAN, but using fewer channels. The latest version of Bluetooth for positioning is known as Bluetooth low energy (BLE) [81]. Bluetooth beacons determine the distance between the two nodes using the following methods: angle of arrival (AoA), time of arrival (ToA), fingerprinting, and received signal strength indicator(RSSI). These methods are explained in Section 2.1.3. Due to the complexity of implementation and to keep costs low, most of the Bluetooth positioning is done using the RSSI-based method. As the BLE name suggests, the beacons consume less power. With low cost and low power requirement, there have been a number of commercial implementations in the market including iBeacon, Eddystone, and piBeacons [82, 83]. Although the beacons are inexpensive and energy

efficient, their positioning accuracy is low (less than 2 m for 40% of the estimated values) [84]. Accuracy issues have led to discontinuation of some of the beacons such as Eddystone. Since accuracy is of a primary importance for our application, Bluetooth beacons were not considered.

2.1.7 Ultrasonic

Ultrasonic localisation uses ToF or TDoA sound waves with frequencies higher than the human audible range (more than 20 kHz), to measure the distance from single or multiple transmitters to single or multiple receivers. Most of the accurate ultrasonic positioning systems developed in past research, such as Marvelmind [85], Active Bat[86], or Cricket [87], used ToF with trilateration to determine the location of an object within a few centimetres accuracy. Trilateration is illustrated in Figure 2.8. Typically, the ultrasound signal is transmitted along with the radio frequency signal to synchronise the transmitter and the receiver for the TOF calculation. Despite the positioning accuracy achieved in past research, ultrasonic signals, unlike radio frequency, depend on the medium of transmission (air) which varies with change in humidity, temperature, and when there are other sources of interference in the environment. This reduces the positioning accuracy of the system. Also, similar to UWB, ultrasonic systems are high-cost, and require a power source at both transmitter and receiver end, which makes them unviable for our application.

2.1.8 Visible Light Communication

Visible Light Communication (VLC) uses modulated visible light signals (between 385 THz to 800 THz) emitted by light emitting diodes (LEDs) to communicate with receivers (photodiodes or camera). Similar to WLAN, the communication signals are used for positioning using either the RSSI or AoA methods. Among the two, AoA is the most accurate and more commonly used, and it can achieve 5-10 cm level of accuracy [88, 89]. Even the RSSI-based implementation with trilateration has attained a positioning error of 40 cm [90]. Whilst VLC is highly accurate, LEDs are expensive to manufacture, as modulation circuitry involves clock and control modules which increase the cost of the localisation solution. Additionally, LEDs have higher power consumption (around

3 Watts an LED [91]) as they are primarily designed for illuminating the space. High cost and power consumption make them unfeasible for our application.

2.1.9 Digital Differential Barometer

The differential barometer system uses relative atmospheric pressure to measure height or detect change in floor levels. This system usually consists of at least two digital barometers: one for the reference, and the other to measure the change in pressure. Both sensors measure the absolute atmospheric pressure which changes with the altitude; higher altitudes have lower atmospheric pressure and vice versa. This can be used to measure the displacement in vertical direction or a change in floor level in a building. Even though the barometer measures the absolute atmospheric pressure at a particular location or a floor level, the same value is not reliably reproducible at a later time due to the change is weather condition or even wind [92]. This prevents them being used as an absolute measurement device. To resolve this, researchers have used a referenced barometer in the building, usually mounted at ground level. The change in height or floor level is determined by comparing the measured pressure value to the reference pressure value [93, 94]. An example implementation of a differential barometer by Kim et al. is shown in Figure 2.9 [95].

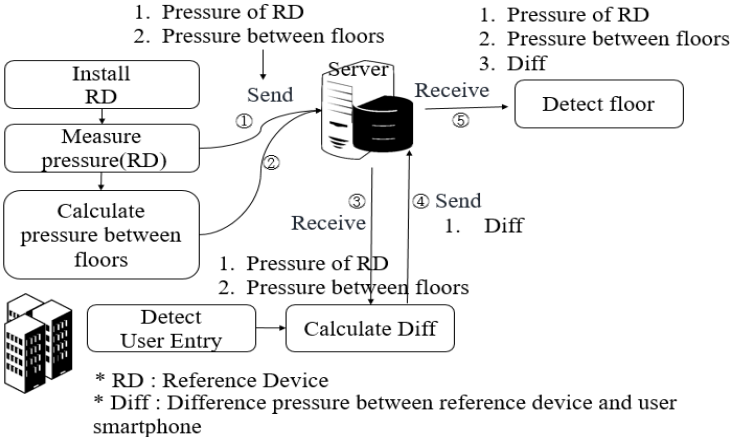


FIGURE 2.9: Differential barometer implementation [95]

2.1.10 Magnetometer

A magnetometer measures the orientation of a device by using the earth's magnetic field. The measured orientation is in the horizontal plane around the z-axis, and gives absolute

orientation information from the global frame of reference. The earth's magnetic field is relatively weak, so magnetometers are easily effected by a magnetic field produced from nearby magnetic objects. Therefore, in the presence of an external magnetic field they are not a reliable source of orientation information. In research conducted by Rai et al., orientation errors of up to 15° were found 90% of the time in a large office [96]. In other work by Afzal et al., orientation errors of up to 100° were observed in an office building, a school's student centre, and a mall [97], and they concluded that different building materials have different effects on the magnetometer.

A magnetometer's data can be fused with gyroscope data to correct the drift and improve the overall orientation accuracy [98]. The two sensors have complementary error characteristics: gyroscopes perform poorly over a long-term orientation computation due to the drift error accumulation, whereas magnetometers are prone to short-term orientation errors due to magnetic field disturbances. Therefore, a sensor fusion between the two can achieve a better overall orientation solution in outdoor environments, which are less prone to magnetic field disturbances [55]. Magnetic field disturbance can also be reduced by considering the magnetic material causing the disturbance [99, 100] and calibrating the device for the disturbance.

2.2 Comparative Analysis of the Technologies

In past research, there have been several performance parameters used to compare different position technologies, such as localisation techniques and methods (RSS/trilateration, ToA/trilateration, TDoA/trilateration, AoA/triangulation etc.), accuracy, range, power consumption, infrastructure, scalibility, privacy, security, and cost [3, 101]. Among these, there are five parameters that are important for our application: accuracy, range area, power consumption, cost, and infrastructure requirements, and they are considered for the comparison. Table 2.1 summarises the performance of 10 different technologies which we considered for our survey against these parameters.

The accuracy parameter in the table refers to the Euclidean distance error between the estimated and the ground truth position. Range is the maximum distance between the mobile agent and a landmark when the position value of the agent can still be calculated. The power consumption parameter is the energy consumed by the positioning

system over time. Most past research work has neglected the power consumption parameter; therefore, the power consumption values have been approximated from the system description in the referenced works. In the table, the power consumption has been categorised into three groups; high (more than 2 Watts), medium (less than 2 watts but more than 500 milliwatts), and low (less than 500 milliwatts). Similar to power consumption, cost of the technology is also estimated based on the system description and categorised into three groups; high (more than $300 USD), medium (less than $300 USD but more than $50 USD), and low (less than $50 USD). Finally, the infrastructure parameter is classified as active if it requires a power source or passive when no power source is needed.

TABLE 2.1: Comparison of the indoor positioning technologies.

Technology	Accuracy	Range	Power Consumption	Infrastructure	Cost	Limitations
Camera	6 cm – 2.5 m	1 m – 10 m	High	No / Passive[a]	Low	Dependence on lighting conditions
WLAN	1.5 m – 5 m	20 m – 50 m	High	Active	Medium	Signals are affected by the changes in the environment
UWB	15 cm – 1.5 m	10 m – 20 m	High	Active	High	Highly accurate clock requirement
RFID	3 cm – 5 m	5 cm – 30 m	High	Active / Passive[b]	High	Short range of the passive tags, and interference from other RFID readers
Bluetooth	2 m – 8 m	5 m – 10 m	Low	Active	Low	Weak connection between devices
Ultrasound	2 cm	2 m – 10 m	Medium	Active	High	Medium of transmission
VLC	10 cm – 40 cm	1 m – 10 m	High	Active	High	Interference from other light sources
Inertial Sensors	2 m – 5 m	Unlimited	Low	No	Low	Accumulates error
Barometer	3 m – 10 m	Unlimited	Low	No	Low	Only measures change in height and gets effected by weather variation
Compass	2 m	Unlimited	Low	No	Low	Hard and soft iron effect

[a] Camera-based systems using natural feature points in the environment do not require infrastructure, whereas ones using fiducial markers need passive infrastructure

[b] RFID tags could be battery powered, active tags, or passively powered, passive tags

2.2.1 Discussion:

Among the main indoor positioning technologies used for localising humans in large-scale indoor environments, UWB, ultrasound, VLC, active RFID and camera-based localisation technologies are the most accurate positioning systems. UWB, active RFID, VLC power consumption at the infrastructure is quite large, and require supply from the

electrical grid for long-term positioning system which makes them unviable for our application. Ultrasonic transmitters consume less power; however, unlike radio frequency, the ultrasonic signals depend on the transmission medium (air) that leads to change in accuracy with variation in temperature or humidity. Also, increasing the number of transmitters increases the signal interference, which hinders the scalability of the system. UWB, ultrasound, VLC, and active RFID sensor devices are also high-cost, making the installation over larger areas prohibitively expensive.

Compared to other accurate positioning systems, camera-based localisation is low-cost, as camera and fiducial markers are the only hardware required. Cost of a full high definition camera is around $40 USD, whereas printing a fiducial marker costs a few cents. Although camera-based localisation gives a higher accuracy and low cost solution, it consumes a lot of power and depends on external lighting conditions. Camera-based localisation systems can be adapted to external lighting by the choice of fiducial markers and installing external lights, however, power consumption still remains a limitation. This can be addressed by combining another technology which is low-power consuming, low-cost, and does not require active infrastructure. Inertial sensors fulfil these requirements, and it was selected along with the camera-based localisation for the developed positioning system.

2.3 Sensor Fusion Approach

Each sensor used in a positioning system has an inherent uncertainty which decreases the overall accuracy of the system; however, the accuracy can be improved by integrating complementary sensors. The fusion of the sensors is achieved with Bayesian filters, which provide a probabilistic framework to combine the sensors. Bayesian filters consider the individual uncertainty of each sensor to estimate the state of the positioning system. The state of the positioning system captures the position, orientation, velocity, and sensors' errors. To estimate the current state x_t of the system Bayesian filter requires all the previous measurements made up to time t. This increases the computational complexity as more measurements are made, so to make the computation more manageable a Markov assumption is made. In the Markov assumption, the current state estimation depends only on the previous state and not the entire history, and the system is described by two models: a motion or propagation model which defines how the system's state changes

with time, and a measurement model which uses sensor measurement to correct the propagated state. A positioning system can have multiple sensors for measurement correction, with each sensor having a different measurement model. Bayesian filters provide a probabilistic framework to combine these measurement models from different sensor sources; this allows more accurate state estimation of the system. There are a variety of methods to implement the Bayesian filter, and each method has different levels of computational complexity and accuracy output. The most common methods in the literature are the Kalman filter and the particle filter. These sensor fusion methods can be applied in a loosely or tightly coupled configuration; we will also discuss the two configurations in the context of visual-inertial sensor fusion.

2.3.1 Kalman Filter

The Kalman filter is the most popular Bayesian filter of all the variants, as it is the most efficient computationally. The Kalman filter assumes that the positioning system is linear, and the system state value is Gaussian distributed which means it can be represented by a mean and a covariance. The state value is tracked over time by tracking the mean and covariance. The Kalman filter is a Bayes filter with Markov assumption, so it can be described by two models: a motion model, and a measurement model. The Kalman filter motion model and measurement model inputs and outputs are illustrated in Figure 2.10. The Kalman filter assumes a linear motion and measurement model. The linear motion model estimates the state x_t from the previous state $x_{(t-1)}$ and control vector input $u_{(t-1)}$ and is given by [102]:

$$x_t = Ax_{(t-1)} + Bu_{(t-1)} + w_{(t-1)}, \tag{2.2}$$

where A and B are matrices with constant value, and $w_{(t-1)}$ is the process noise. Covariance X_t of the state from motion model is given by:

$$P_t = AP_{(t-1)}A^T + Q, \tag{2.3}$$

where Q is the covariance of process noise, and it is also a Gaussian distribution. The Measurement model is given by:

$$z_t = Hx_t + v_t, \tag{2.4}$$

and updates the state x_t using the sensor measurement value z_t. H is the matrix in the equation that defines the relation with current state x_t, and v_t is the measurement noise. Covariance Z_t of the measurement model is given by:

$$Z_t = HP_tH^T + R, \tag{2.5}$$

where R is the covariance of the measurement noise. The new state and covariance are calculated after the measurement update by first computing the Kalman gain, K_t:

$$K_t = P_tH^T(HP_tH^T + R)^{-1}, \tag{2.6}$$

then the state:

$$x_{(t+1)} = x_t + K_t(z_t - Hx_t), \tag{2.7}$$

and the covariance:

$$P_{(t+1)} = (I - K_tH)P_t \tag{2.8}$$

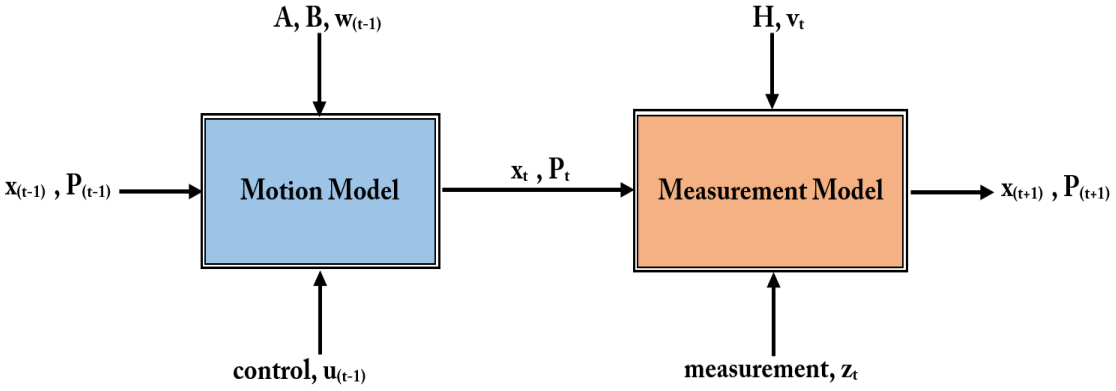


FIGURE 2.10: Kalman filter inputs and outputs

The Kalman filter described above can only be applied to linear systems, but most of the real world systems including the positioning system are non-linear. However, a Kalman filter solution can be extended to approximate non-linear systems using the extended Kalman filter (EKF). The EKF linearises the non-linear motion and measurement model using the tangent or Jacobian at time t. In EKF, the motion model's A and B matrices are replaced by a function f_t and the motion model is given by:

$$; x_t = f_t(x_{(t-1)}, u_{(t-1)}) + w_{(t-1)} \tag{2.9}$$

For the measurement model, H is a function of the current state x_t, so Hx_t in Equation 2.5 is replaced by $h_t(x_t)$ resulting in:

$$z_t = h_t(x_t) + v_t \tag{2.10}$$

The Jacobians of the function f_t and h_t are given by:

$$F_t = \frac{\partial f}{\partial x}|x = x_t, \tag{2.11}$$

and

$$H_t = \frac{\partial h}{\partial x}|x = x_t, \tag{2.12}$$

respectively. Even though EKF can be applied to a non-linear system, it still assumes that the state is Gaussian distributed. The state in our case is position and heading values of a first responder, which mostly remain Gaussian. However, sometimes the values will deviate from Gaussian distribution. This could lead to inaccurate state estimation compared to the particle filter, which is discussed in the next section. Although, EKF is not as accurate as a particle filter, it requires a lot less memory and computational resources which in turn results in less power consumption as well. Since the focus of this thesis is energy efficiency and accuracy, EKF was chosen for our developed system.

2.3.2 Particle Filter

The particle filter (PF) is a Bayesian filter that can be applied on a non-linear system. It approximates the state of the system by using a set of particles where each particle carries a weight depending on its likelihood to represent the true value of the state [103]. Similar to other Bayes filters, the PF also calculates the current state using the previous state value. There are different variants of PF, but they all perform these four sequential steps: initialisation, propagation, correction, and resampling [104]. First, during initialisation, a finite set of particles is created to represent an initial state x_{t_0}, and they are initialised with unity weight. Then these particles are propagated based on the the system's motion or propagation model. In the third step, the particles' weight values are corrected based on their proximity to the measured value. Particles which are closer to the measured value z_t result in a higher weight w_t value. Finally, the particles are resampled based on their newly calculated weights to create new copies of

the existing particles with equal normalised weight. The next iteration of the filter starts from the propagation step and ends at resampling, and the cycle continues recursively. The set of particles at time t can be represented mathematically by $\{x_t^i, w_t^i\}_{i=1}^N$ where N is the number of particles, x_t^i is the state of the ith particle and w_t^i is the associated weight, and the value of i ranges from 1 to N [105]. The main advantage of the PF is the accurate estimation of the state value in a multi-modal non-linear state space. However, in order to have more accurate state estimation or localisation results, a larger number of particles is required, which results in higher computation, memory, and energy cost [106–109]. This is also the reason why the PF was not chosen for our developed system.

2.3.3 Visual-Inertial Sensor Fusion

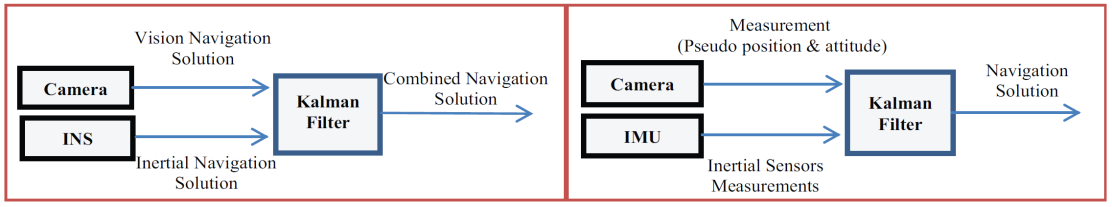
The two sensors, camera and inertial measurement unit (IMU), have complementary characteristics. The camera is an exteroceptive sensor that captures information in the environment and uses it to estimate the pose (position and orientation). If there is a lot of information (rich textures) in the environment, the camera can accurately estimate pose over the long-term. In contrast, IMU is a proprioceptive sensor that internally measures acceleration and angular velocity and uses the measured values to calculate pose. IMUs are accurate for short-term pose estimation but are inaccurate over the long-term due to the drift accumulation. Additionally, a camera captures data at a low frame rate (approximately 30 fps on average) with high power consumption (more than 1.5 watts) and computational cost, whereas IMU has a high frame rate (nearly 100 Hz or more) with low power consumption (30 to 40 milliwatts) and computational cost. A sensor fusion between the camera and IMU can be used to optimise the pose accuracy and power consumption of the localisation system.

Visual-inertial localisation methods can be categorised into visual-inertial odometry (VIO) [31, 110–113] or visual-inertial simultaneous localisation and mapping (VISLAM) [114–118] approaches. VIO algorithms use visual features to estimate camera pose and combine it with inertial pose using motion constraint. Each VIO pose estimate is with respect to the last pose estimate that is joined together to give motion trajectory. Since there is no global reference information (i.e., previously visited location, also known as loop-closure), VIO algorithms are primarily dead reckoning (odometry) approaches. By contrast, VISLAM methods simultaneously use visual features to localise and create a

map for future reference. If a visual feature in the current frame is identified to be in the priori map (global reference), it is also used to calculate and correct the current pose using loop-closure. There are mainly two configurations to fuse the visual and inertial sensor data: loosely or tightly coupled configuration, which are discussed below.

a) Loosely-Coupled Fusion

In loosely-coupled fusion (illustrated in Figure 2.11 a), visual and inertial sensors estimate the pose separately and combine the pose values using motion constraints [111, 119–122]. This approach is more suitable for non-rigidly attached sensors (for example, a head-mounted camera combined with a foot-mounted IMU estimating the 2D pose of a person). Loosely-coupled sensor fusion methods can also be used for rigidly attached sensors to decrease the computational cost; however, this results in a less accurate localisation solution.



(a) Loosely Coupled Integration (b) Tightly Coupled Integration

FIGURE 2.11: Visual-inertial sensor fusion configurations [123]

b) Tightly-Coupled Fusion

The tightly-coupled sensor fusion (illustrated in Figure 2.11 b) directly fuses the visual and inertial sensors' data, and it requires the sensors to be rigidly attached. This technique cannot be applied to non-rigidly attached sensors, which is the case for our system. Tightly-coupled sensor fusion methods lead to higher accuracy pose estimates but require more computational resources [31, 110, 117, 124, 125].

Past research on VIO and VISLAM has indicated that they can provide accurate pose estimation for small-scale environments, but they are not robust for long-term, large-scale, and safety-critical localisation [126–128]. Additionally, visual features essential for the camera pose estimation may be sparsely present in certain environments, such as training environments made-up of shipping containers or indoor spaces with bland textureless walls. In these environments, placing fiducial markers to provide a global location source improves the consistency and robustness of the visual-inertial localisation solution.

2.4 Conclusion

In the first part of this chapter, indoor positioning technologies for localising first responders were reviewed. Each technology was analysed with its advantages and disadvantages, and potential application to our system. A comparative analysis was also done using the performance parameters required for the targeted application. The comparison is summarised in Table 2.1. From the comparison, UWB and radio signal-based positioning methods provide a more accurate positioning solution; however, these systems require a power source at the infrastructure and are expensive to install. From the analysis, two technologies that offer accurate, energy efficient, low-cost, and large-scale indoor positioning solution emerged:

- **Inertial Sensors:** Inertial sensors are a low-cost and energy efficient positioning technology which uses an internally referenced mechanism to estimate localisation information. Inertial sensors do not require any infrastructure installation and are not limited by range. However, their position and orientation estimate accuracy deteriorates after some time due to noises and biases in the sensors, and also because of reliance on previous values for update. Therefore, it requires a correction mechanism to reduce the accumulated errors.

- **Camera-based Localisation System:** Similar to inertial sensors, camera-based localisation is a low-cost technology, but unlike inertial sensors, can calculate position and orientation values accurately long-term. However, the camera needs an external frame of reference of feature points to determine the position and orientation values. Feature points are generally insufficient in indoor environments such as training environments or office spaces, so to add these, fiducial markers, with their low-cost, passive infrastructure, can be installed. Power consumption of camera-based localisation solutions are typically medium to high, as camera sensor and electronics require a comparatively larger amount of energy to operate. The range of a camera-based localisation system varies from 1 to 10 m depending on the resolution and the size of fiducial markers.

In the second part of the chapter, sensor fusion algorithms to combine different positioning technologies were discussed. To have more in-depth understanding of these methods,

the implementation details of the sensor fusion algorithms were also described. Strengths and weaknesses of different sensor fusion algorithms were also examined and are summarised as follows:

- **Linear Kalman Filter** is the most efficient method of combining sensors. It requires fewer computational resources to implement. However, it is only applicable to linear systems.
- **Extended Kalman Filter** is derived from the linear Kalman filter; it keeps some of its characteristics such as efficient implementation, and computationally requires fewer resources. Unlike the linear Kalman filter, it can also be applied to non-linear systems, which is the case for our system. However, it still assumes that the state space is Gaussian distributed.
- **Particle Filter** can be used with non-linear systems, and is also applicable to non-Gaussian state spaces. It is more accurate than the Kalman filters; however, it requires a lot of computation, memory and power resources, and the solution typically needs more time to complete.

From these methods, the extended Kalman filter was chosen for the developed system, as it is the most efficient and resource conserving implementation, and is also applicable to non-linear motion and measurement models.

The next chapter goes into details of the chosen positioning technologies: the inertial sensor and camera-based localisation system. It discusses the underlying coordinate systems to compute position and orientation values from these technologies, and the transformation between the coordinate systems. It also analyses the errors present in these technologies and suggests ways to mitigate the errors.

Chapter 3

Fundamentals of the Inertial and Camera-based Localisation

The localisation system developed for this thesis uses inertial and visual sensors to estimate position and orientation values. This chapter examines the fundamentals of these sensors, including the coordinate systems involved (Section 3.1), the sources of errors in inertial sensors and their mitigation (Section 3.2), and the mathematical transformations required in visual sensors to compute the localisation values (Section 3.3). Lastly, the chapter discusses the visual sensor calibration method which has been used to compute the intrinsic parameters of the sensor (Section 3.3.2).

3.1 Coordinate Frames

Localisation system measures position, velocity, and orientation as vectors relative to a known location or reference frame [129]. Understanding these reference frames is essential to relate the measured sensor values to the computed localisation information. There are several reference Cartesian coordinate frames which describe the navigation state (position, velocity, and orientation) of a moving body with reference to its environment. A basic localisation system consists of the following coordinate frames: sensor frame, body frame, Earth-centred inertial frame, Earth-centred Earth-fixed, and navigation frame.

3.1.1 Sensor Frame

The sensor frame (s-frame) is the reference coordinate system of the inertial or visual sensor attached to a moving body to measure the navigation state of the body. Figure 3.1 illustrates the s-frame for the inertial and visual sensors. The axes of the s-frame are aligned with the orientation of the inertial and visual sensors. The origin is a corner point of the MEMS chip for the inertial sensor and the optical centre for the camera.

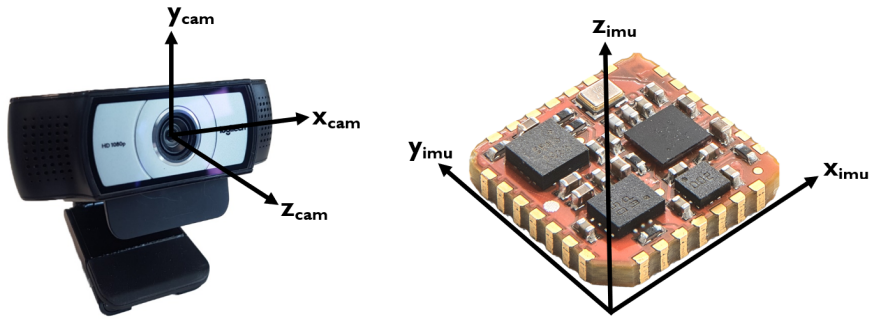


FIGURE 3.1: Sensor frames for visual (left) and inertial (right) sensors

3.1.2 Body Frame

The body frame (b-frame) is the coordinate system associated with the moving body. Its origin is located at the centre of mass of the body. The y-axis, x-axis, and z-axis of the b-frame points towards the forward, sideways, and upward direction of the body, respectively. Figure 3.2 illustrates the b-frame as well as the relationship of the s-frames (inertial and visual sensors) to the b-frame.

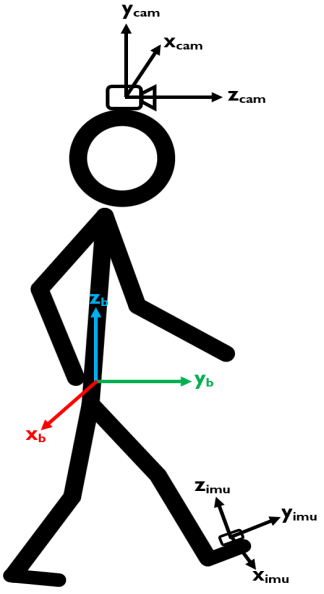


FIGURE 3.2: Body and sensor frames for a moving body

3.1.3 Earth-Centred Inertial Frame

The Earth-centred inertial frame (i-frame) is the inertial coordinate system where Newton's laws of motion apply, and the Earth is assumed to be stationary. The i-frame is illustrated in Figure 3.3 along with its relationship to other frames. The i-frame's origin is at the centre of the Earth. The z-axis of the i-frame is along the Earth's rotational axis. The Equator is the xy plane, where the x-axis points towards the vernal equinox, and the y-axis is given by the right-hand rule.

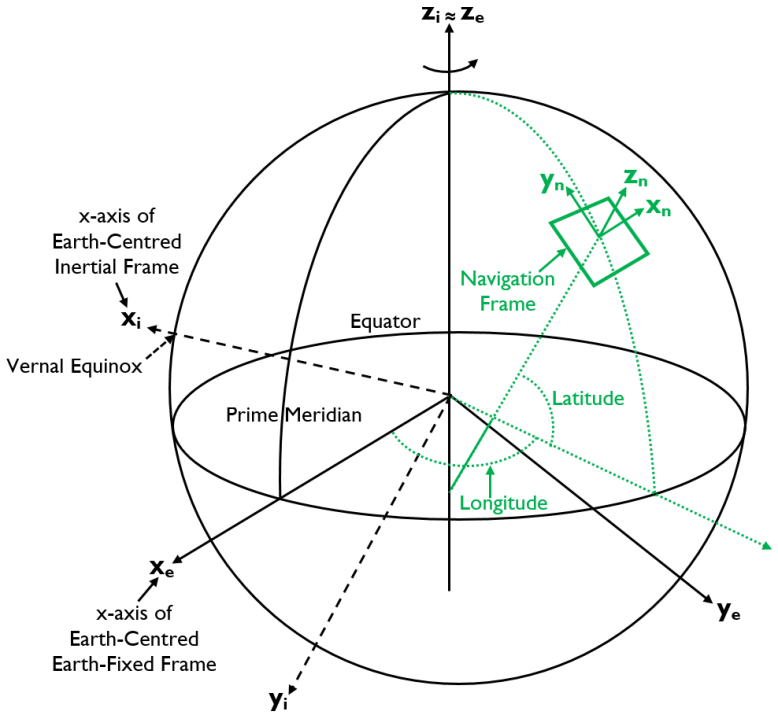


FIGURE 3.3: Earth-centred Inertial (x_i, y_i, z_i), Earth-centred Earth-fixed (x_e, y_e, z_e), and Navigation (x_n, y_n, z_n) frames and their relationship to each other.

3.1.4 Earth-Centred Earth-Fixed Frame

The Earth-centred Earth-fixed frame (e-frame) is the geodetic coordinate system defined by the World Geodetic System 84 (WGS84) datum [130], and it provides a reference to all locations on the Earth. The e-frame is illustrated in Figure 3.3 along with its relationship to other frames. The e-frame rotates with the Earth and is described in terms of longitude and latitude angles measured with respect to the equatorial plane and the prime meridian respectively. The z-axis of the e-frame is along the Earth's rotational axis; the x-axis is on the equatorial plane and points towards the prime meridian, and the y-axis is derived by the right-hand rule.

3.1.5 Navigation Frame

The navigation frame (n-frame) is the local coordinate system which is fixed and tangential to the Earth's surface. The origin of the n-frame is located at the centre of mass of the moving body or at an arbitrary point fixed on Earth's surface. The z-axis of the n-frame points upward, normal to the Earth's ellipsoid. The xy-plane is tangential to the Earth's surface where the x-axis points east, and the y-axis points north. Figure 3.3 illustrates the n-frame and its relationship to other frames. In this thesis, we made two assumptions on the navigational frames: first, that the distance travelled by the moving body (walking first responders) is relatively short (few kilometres), which reduces the effect of the curvature of the Earth to a negligible level; and second, that the consumer-grade low-cost inertial sensors do not detect the Earth's rotation [131]. With these two assumptions, i-frame, e-frame, and n-frame are reduced to one reference frame, n-frame. Therefore, only three reference frames are used in this thesis: s-frame, b-frame, and n-frame.

3.2 Inertial Sensors

Inertial sensors have been introduced earlier in Section 2.1.2. Different types of inertial sensors and the methods to compute the localisation information from their data have also been discussed in Section 2.1.2. This section covers the error characteristics of the inertial sensors and the methods to reduce these errors.

3.2.1 Errors in Inertial Sensors

Practical inertial sensors output acceleration and angular velocity different to the true acceleration and angular velocity experienced by the sensors. The difference between the input and output values is the error in measurement. There are two types of errors present in the inertial sensors data: systematic and stochastic errors [132].

a) Systematic Errors

Systematic errors, as the name suggests, can be modelled and mathematically offset in the raw sensor output data through calibration. There are mainly four systematic errors present in the inertial sensor data: bias, scale factor, non-linearity, and cross-coupling [132, 133].

Bias in inertial sensors occurs as a constant non-zero output value when the sensor is in a static state and not experiencing any external perturbation. It is also referred to as *zero-g offset* for an accelerometer and *zero-rate offset* for a gyroscope [26–29]. The units for the bias are the same as the sensor output, deg/s for gyroscope and milli-g (mg) for accelerometer. The bias could be caused by variation in ambient temperature, and it can be corrected through thermal calibration [134]. The graph in Figure 3.4 illustrates the constant shift in the sensor output value due to bias error.

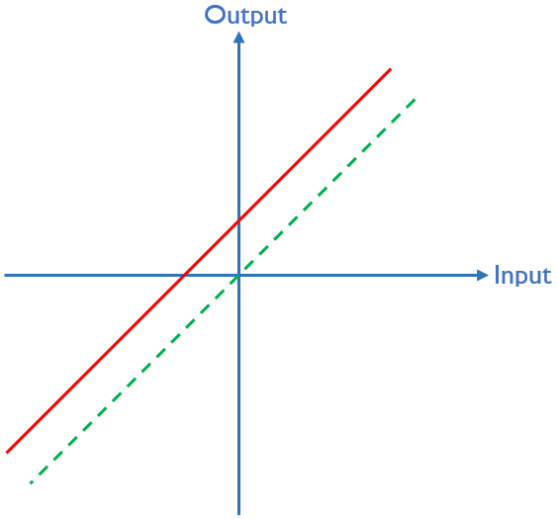


FIGURE 3.4: Bias error [133].

Scale Factor in inertial sensors shows as a ratio by which the sensor output value is scaled up or down from the input value. It is caused when the conversion from sensors' output to physical units is incorrect. The scale factor is also referred to as *sensitivity tolerance* in inertial sensors datasheets and described in % value, for example, 97% or 102% of the input value [26–29]. The graph in Figure 3.5 illustrates the change in the gradient from unity to a ratio due to the scale factor error.

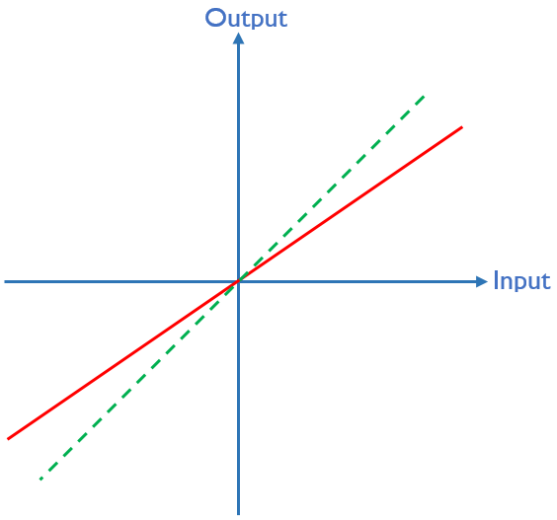


FIGURE 3.5: Scale factor error[133].

Non-linearity in inertial sensors occurs when the output value of the sensor varies non-linearly with the applied input value. The non-linearity is usually described as a % value of full-scale output. In general, the non-linearity shows up near the boundary range of the output [135]. The graph in Figure 3.6 illustrates the effect of non-linearity on the sensor output value.

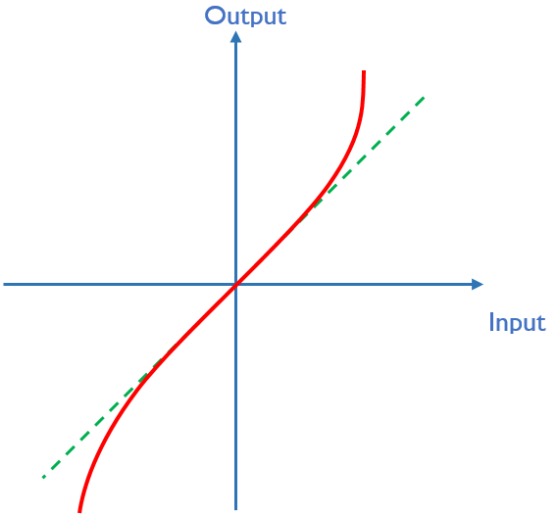


FIGURE 3.6: Non-linearity error [133].

Cross-coupling occurs when the axes of the sensors are physically misaligned and are not fully orthogonal as expected of the ideal sensor. This leads to acceleration or angular velocity on one axis being projected on another axis. Similar to the scale factor, the cross-coupling is described in fix scalar % value. Cross-coupling is also referred to as *cross-axis sensitivity* or *non-orthogonal error* [26–29]. A cross-coupling error is illustrated in Figure 3.7.

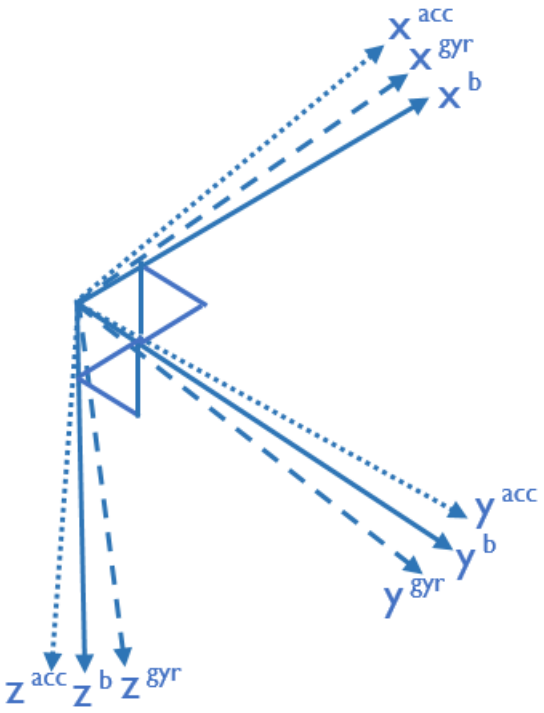


FIGURE 3.7: Cross-coupling error [129].

Temperature-Dependent Bias/Scale Factor shows as a variation of bias and scale factor error with the change in ambient temperature. The error can be very noticeable for low-cost MEMs inertial sensors, and it should be accounted especially in applications which require inertial sensors to operate over multiple days. The unit of bias are: deg/s for gyroscope and milli-g (mg) for accelerometer, and of scale factor are in % value. The temperature-dependent variations can be calibrated by the manufacturer and specified in datasheet, as is the case for XSens MTw Awinda [29], or it should be performed through thermal calibration procedures that involve measuring change in sensor values (acceleration or angular velocity) with the change in temperature [136, 137].

b) Stochastic Errors

Stochastic errors are random in nature, and therefore cannot be estimated and systematically removed through calibration as there is no correlation between the past and future values. However, the values of these errors can be stochastically modelled and their effect on sensor accuracy can be reduced. There are two main stochastic errors present in the inertial sensor data: white sensor noise and bias instability [132, 133].

White Sensor Noise is a random uncorrelated zero-mean error in inertial sensor value which is caused by thermo-mechanical perturbations. White sensor noise is integrated and usually specified as an angular random walk or velocity random walk for gyroscope

and accelerometer, respectively. The units for an angular random walk and velocity random walk are $deg/\sqrt{Hz}$ and micro-g$(\mu g)/\sqrt{Hz}$, respectively. White noise is also referred to as *random noise* or *noise* in inertial sensor datasheets [26–29].

Bias Instability is a measure of how stable the inertial sensor bias is over a time period. A lower bias instability value implies higher stability and less random drift in bias value, which is advantageous. Bias instability is specified in deg/h or deg/s, and it is also referred to as *bias drift* or *in-run bias stability* in inertial sensor datasheets [26–29].

3.2.2 Methods to Reduce Errors

To mitigate these errors, outputs of the inertial sensors are formulated into a mathematical model which is used to calibrate the sensors. The Error model for an accelerometer is given by:

$$\begin{bmatrix} a'_x \\ a'_y \\ a'_z \end{bmatrix} = b_a + S_a \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} + M_a \begin{bmatrix} a_x \\ a_y \\ a_z \end{bmatrix} + n_a, \tag{3.1}$$

and the components of the error model equation are elaborated:

$$b_a = \begin{bmatrix} b_{ax} \\ b_{ay} \\ b_{az} \end{bmatrix}, \; S_a = \begin{bmatrix} s_{ax} & 0 & 0 \\ 0 & s_{ay} & 0 \\ 0 & 0 & s_{az} \end{bmatrix}, \; M_a = \begin{bmatrix} 1 & m_{a,xy} & m_{a,xz} \\ m_{a,yx} & 1 & m_{a,yz} \\ m_{a,zx} & m_{a,zy} & 1 \end{bmatrix}, \; n_a = \begin{bmatrix} n_{ax} \\ n_{ay} \\ n_{az} \end{bmatrix}, \tag{3.2}$$

where (a'_x, a'_y, a'_z) and (a_x, a_y, a_z) are the recorded and the actual acceleration value. b_a is the constant accelerometer bias vector, and n_a is the accelerometer's white noise vector. Matrices S_a and M_a are the accelerometer scale factor and the misalignment, respectively.

Similarly, for the gyroscope the errors are mathematically modelled by:

$$\begin{bmatrix} g'_x \\ g'_y \\ g'_z \end{bmatrix} = b_g + S_g \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} + M_g \begin{bmatrix} g_x \\ g_y \\ g_z \end{bmatrix} + n_g, \tag{3.3}$$

and the components of the gyroscope's error model equation are elaborated in:

$$b_g = \begin{bmatrix} b_{gx} \\ b_{gy} \\ b_{gz} \end{bmatrix}, \; S_g = \begin{bmatrix} s_{gx} & 0 & 0 \\ 0 & s_{gy} & 0 \\ 0 & 0 & s_{gz} \end{bmatrix}, \; M_g = \begin{bmatrix} 1 & m_{g,xy} & m_{g,xz} \\ m_{g,yx} & 1 & m_{g,yz} \\ m_{g,zx} & m_{g,zy} & 1 \end{bmatrix}, \; n_g = \begin{bmatrix} n_{gx} \\ n_{gy} \\ n_{gz} \end{bmatrix}, \tag{3.4}$$

where (g'_x, g'_y, g'_z) and (g_x, g_y, g_z) denote the recorded and the actual angular velocity values, respectively. b_g is the constant gyroscope bias vector, and n_g is the gyroscope's white noise vector. Matrices S_g and M_g are the gyroscope's scale factor and the misalignment, respectively.

The deterministic (bias, scale factor, misalignment) and stochastic (white noise) errors in Equations 3.1 and 3.3 can be measured through laboratory calibration or user calibration [132]. Laboratory calibration is more accurate as specialist equipment (such as motion simulator, oven, and freezer) are used [138]. However, laboratory calibration is also costly depending on the accuracy of the equipment. Mainly driven by the cost, simpler user methods have been devised to reduce these errors and the cost of calibration. The most widely validated method in past research is a six-position test which involves mounting the inertial sensor in a hollow cuboid box and recording the static accelerometer and gyroscope data on each face of the box [139–143].

For our research we used the XSens MTw Awinda inertial sensor, which comes with a laboratory calibrated scale factor and has misalignment and white noise errors specified in the datasheet [29]. However, the accelerometer and gyroscope biases are not specified in the datasheet, so they were calculated by creating a 3D printed cuboidal box and performing a six-position test. The XSens inertial sensor was mounted inside the box along with a three-axial bubble level to ensure the box was levelled for the six position test. The setup is shown in Figure 3.8, and the direction of the inertial sensor's positions for the test are in Figure 3.9. The calculated biases are in Table 3.1, and they were removed from the raw inertial sensor data recorded for experiments in Chapters 4, 5, and 6. To even further reduce the effect of accelerometer and gyroscope bias due to stochastic instability, 20 seconds of data was recorded before each experiment, with the IMU mounted on a user's foot and in a static state. The calculated accelerometer

and gyroscope biases from these 20 seconds were subtracted from the recorded raw accelerometer and gyroscope data.

TABLE 3.1: XSens inertial sensor's accelerometer and gyroscope biases

Axial Values	**Biases values**	**Standard Deviation**
$b_{ax}\,(mg)$	1.132	0.15
$b_{ay}\,(mg)$	1.145	0.46
$b_{az}\,(mg)$	1.467	0.84
$b_{gx}\,(deg/s)$	2.14	1.24
$b_{gy}\,(deg/s)$	2.52	1.15
$b_{gz}\,(deg/s)$	2.23	1.46

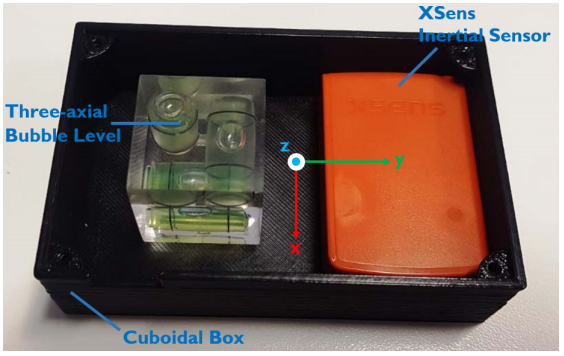


FIGURE 3.8: Six position calibration test setup

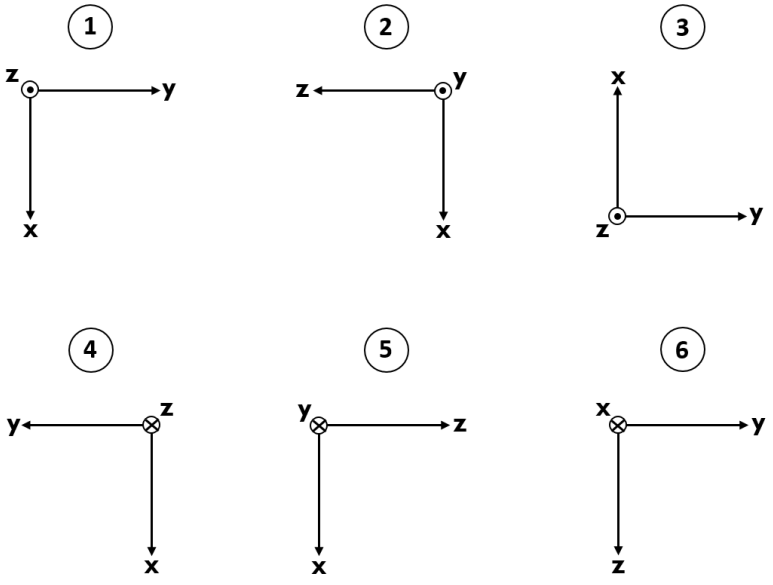


FIGURE 3.9: Inertial sensor and the box orientations for six position calibration. The arrow of each axis indicates its axial direction, a dot in a circle indicates the perpendicular upward axial direction going away from the surface, and a cross in a circle indicates the perpendicular downward axial direction going into the surface

3.3 Visual Sensor

Visual sensors have been briefly discussed in Section 2.1.1 from the application point-of-view. In this section, the fundamentals describing the transformation of a 3D scene to a 2D image by the visual sensor are covered. These are important to understand as they are the key component of our visual sensor's position and orientation estimation method. This section also gives detail on the camera calibration method used for this thesis to compute the intrinsic parameter values of the visual sensor.

3.3.1 Pinhole Camera Model

The pinhole camera model geometrically and mathematically describes the transformation of 3D points in an environment to corresponding 2D image points while assuming that there is no effect of lens distortion on the image [144]. Figure 3.10 illustrates the geometric elements of the pinhole camera model. The model consists of an image plane (also called the *projection plane*), an optical centre (origin of the camera coordinate system), an optical axis, and a principal point (a point on image plane located near the centre at the intersection of the optical axis and the image plane). The optical centre is physically located at the centre of the aperture and the optical axis passes perpendicular to it. Mathematically, the pinhole camera model is defined by intrinsic matrix:

$$M_{int} = \begin{bmatrix} f_x & s_k & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}, \tag{3.5}$$

and extrinsic matrix (M_{ext}):

$$M_{ext} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \end{bmatrix} \tag{3.6}$$

The intrinsic matrix contains the internal parameters of the camera including focal length, (f_x, f_y), the skew factor, s_k, which is distortion caused by non-square pixels, and the principal point of the image plane, (c_x, c_y). The extrinsic matrix is composed of the rotation matrix (representing orientation of the camera with respect to the 3D environment) and the translation vector (representing the position of the camera in the 3D environment). M_{int} and M_{ext} are illustrated in Figure 3.10 and are defined

by Equations 3.5 and 3.6, respectively. The complete transformation of the 3D point $P\,(X_{wi}, Y_{wi}, Z_{wi})$ in the environment to the projected image point $p\,(u_i, v_i)$ using the extrinsic and intrinsic matrix is described by:

$$s \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = M_{int}\, M_{ext} \begin{bmatrix} X_{wi} \\ Y_{wi} \\ Z_{wi} \end{bmatrix} \tag{3.7}$$

Equation 3.7 is also used to compute the intrinsic matrix M_{int} and scale factor s with known points in the real world (such as corner points on a checkerboard) and their projection on the image plane; this process is discussed in the next section. Once the intrinsic matrix is computed, it can be used to compute the unknown extrinsic matrix which contains the localisation information of the camera. Computing the extrinsic matrix for localisation is further explained in Section 4.2.2.

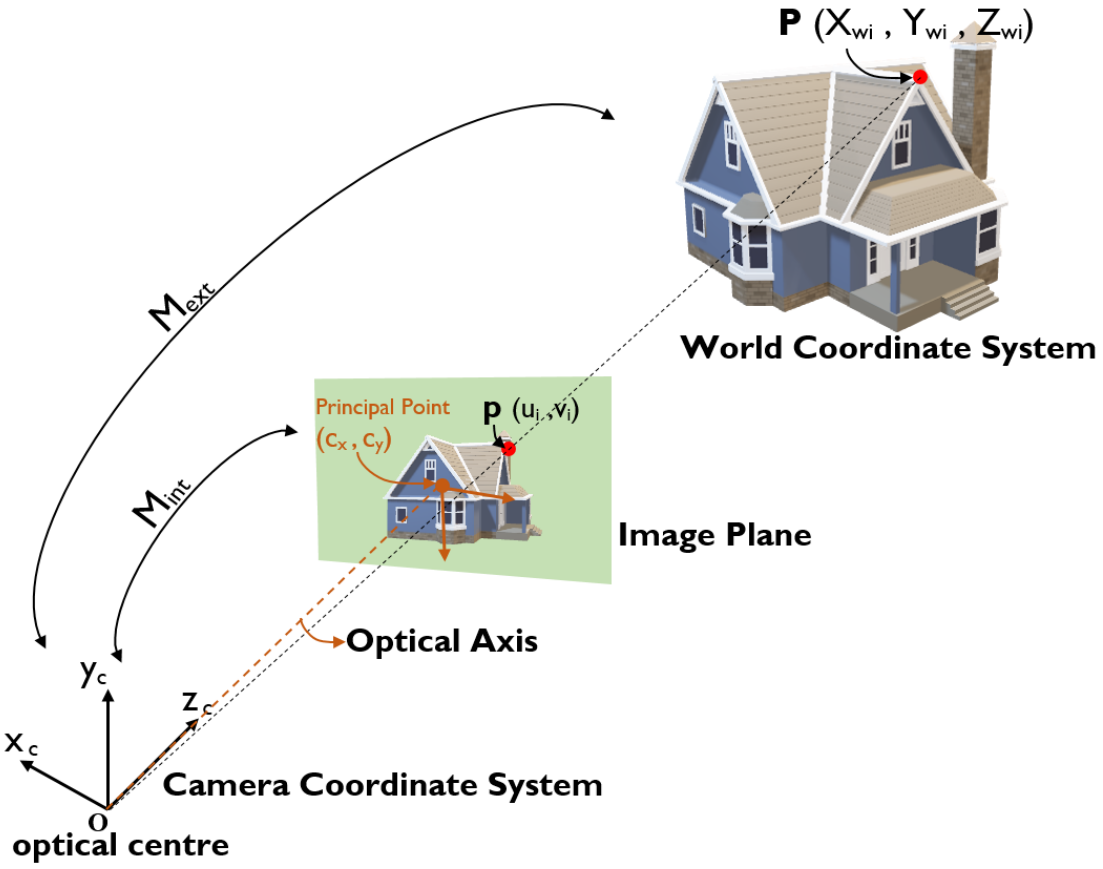


FIGURE 3.10: Pinhole camera model showing the projection of a 3D point to a 2D point in the image plane.

3.3.2 Intrinsic Parameters and Calibration

The intrinsic matrix defines the internal relation of the camera coordinate system and the image plane. Most of the elements in the intrinsic matrix have constant values except for the focal length. The focal length can also be configured to a constant value for web cameras, such as Logitech C920 (the camera used in this thesis) by setting the camera to a manual focus. For high-end cameras, such as photography cameras, the focal length is set to a constant value by using a prime lens. The camera for this thesis was also configured to a fixed focal length with the widest depth of field to keep everything focused in the environment. Once the camera focal length is fixed, all the parameters in the intrinsic matrix have constant values. These parameters only need to be determined once and can be used in later experiments without any modifications. The method to find the intrinsic matrix of the camera is known as camera calibration. During camera calibration, multiple images of a known checkerboard pattern are taken from different viewing angles. The checkerboard pattern contains the known corner points (X_{wi}, Y_{wi}, Z_{wi}), and the images of the checkerboard contain the corresponding image points (u_i, v_i). These two sets of data are used to compute the projection matrix M which is the multiple of intrinsic M_{int} and extrinsic M_{ext} matrices and given by:

$$M = M_{int} * M_{ext} = \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{bmatrix} \tag{3.8}$$

Each corresponding known corner point and image point results in:

$$u_i = \frac{m_{11}X_{wi} + m_{12}Y_{wi} + m_{13}Z_{wi} + m_{14}}{m_{31}X_{wi} + m_{32}Y_{wi} + m_{33}Z_{wi} + m_{34}}, \tag{3.9}$$

and

$$v_i = \frac{m_{21}X_{wi} + m_{22}Y_{wi} + m_{23}Z_{wi} + m_{24}}{m_{31}X_{wi} + m_{32}Y_{wi} + m_{33}Z_{wi} + m_{34}} \tag{3.10}$$

The solution of Equation 3.9 and Equation 3.10 over multiple images gives us the projection matrix which is decomposed to find the intrinsic matrix. We used OpenCV's camera calibration function to implement this process [145].

We performed the camera calibration by using an A3 size checkerboard with 8x5 squares, 7x4 vertices, and squares of size $50mm$ [146]. The checkerboard was mounted on the ceiling of the lab as shown in Figure 3.11, and the Logitech C920 camera was on the helmet of the user. Multiple photos were taken of the checkerboard at different viewing angles. The checkerboard pattern specification and size were passed to the OpenCV function along with the recorded images. The OpenCV camera calibration function calculated the intrinsic matrix values, and the final computed values were used to compute the reprojection error. The mean reprojection error for our camera calibration was 0.278 pixels, which was below the 0.5 pixels recommended by peers [147, 148]. Figure 3.12 shows the 28 (7x4) detected checkerboard vertices and the reprojected points.

Another visual sensor aspect that should be considered for camera pose estimation is the shutter type. There are two shutter types used by the cameras: global and rolling shutter. Global shutter involves capturing an image at a single time instant across the whole image sensor. Whereas for rolling shutter, image gets captured sequentially on the image sensor by each individual row [149]. Majority of the modern cameras use rolling shutter including the Logitech C920 camera used in our localisation system. Due to sequential row by row image capture at the sensor, rolling shutter cameras could introduce artifacts on the captured image if the captured object or camera platform is moving at a high speed. For example, a drone moving through an environment at a high speed, or a camera capturing a fan's blade [150, 151]. The rolling shutter artifacts decrease the localisation accuracy of the system. When the camera platform or the captured object are at a low speed, rolling shutter does not effect the captured image as is the case for our localisation system. However, in the future research when running motion type is investigated, rolling shutter artifacts should be examined in the captured image frames.

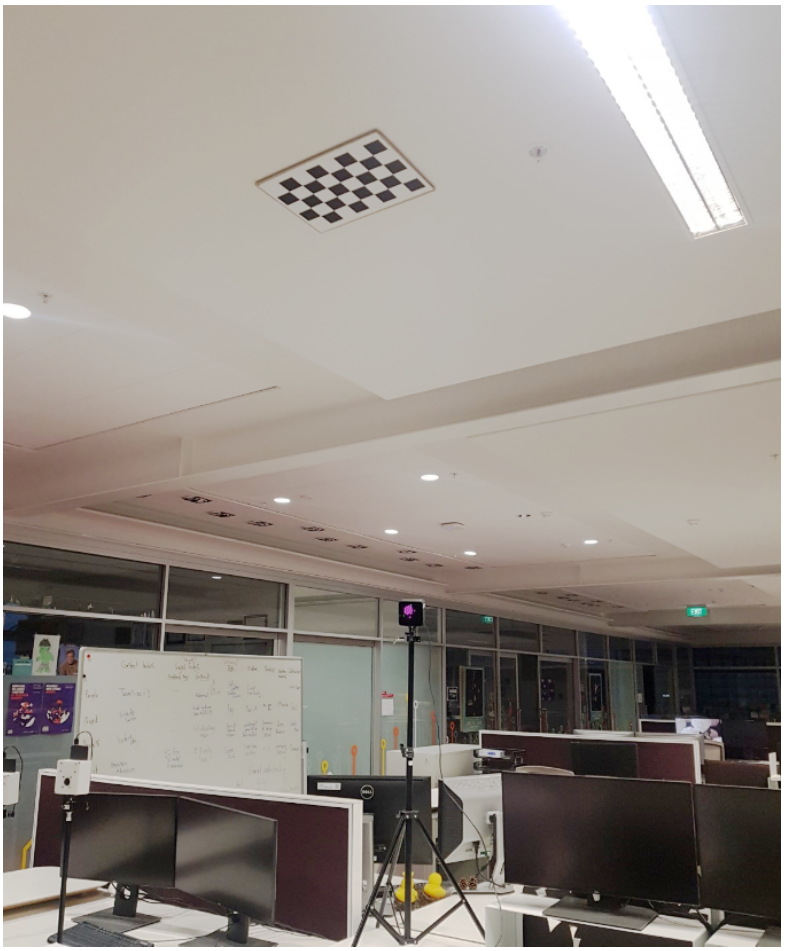


FIGURE 3.11: Ceiling mounted checkerboard used for the camera calibration

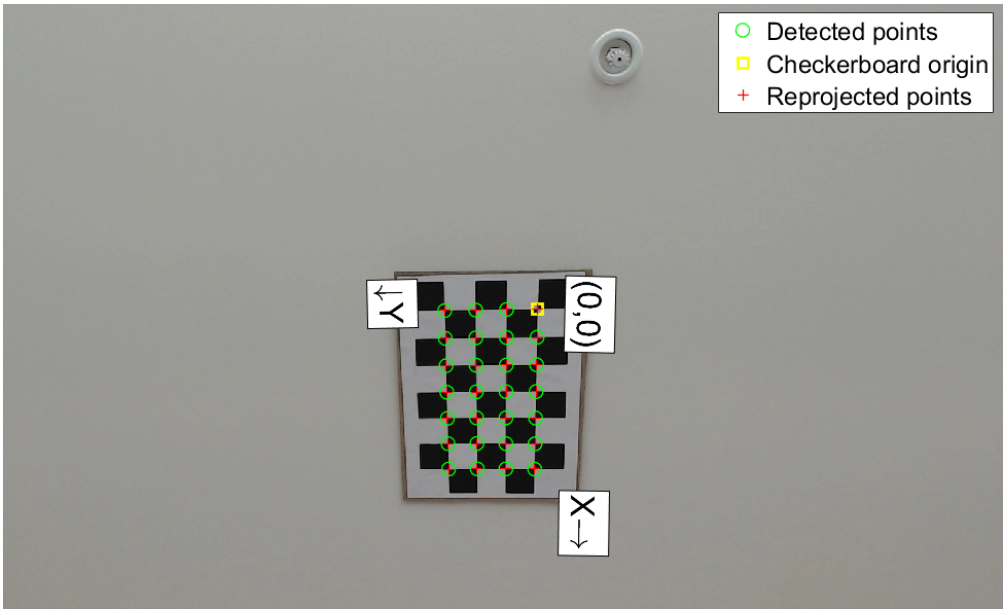


FIGURE 3.12: Detected checkerboard vertices shown by green circles and the reprojected pixels shown by red crosses

3.4 Conclusion

In this chapter, we looked into the fundamentals of inertial and visual sensors. We briefly introduced the coordinate system involved to reference the sensors' data. For inertial sensors, we examined the systematic and stochastic errors present, and the mathematical model used in this thesis to reduce their effect. We also described how these errors were calibrated for XSens IMU, including the six-position test used to estimate a sensor's bias. For the visual sensor, we discussed the pinhole camera model and the mathematical relationship between the 3D point in an environment and its corresponding 2D image point. We also explained the method to calculate the intrinsic parameters of the camera using the mathematical relationship and known points in the 3D environment in the form of a checkerboard. These fundamentals are essential and are used in Chapter 4 to mitigate error in inertial sensors and compute position and orientation values for the camera.

Chapter 4

Loosely Coupled Sensor Fusion Approach

At the end of Chapter 2, we reviewed different sensor fusion approaches: Kalman and particle filters. These approaches can be applied in mainly two different configurations: tightly coupled and loosely coupled [152]. A tightly coupled solution requires sensors to be rigidly attached to each other, and it produces a more accurate localisation solution for drones and wheel-based robots as both the IMU and the camera go through same motion dynamics [152–155]. By contrast, a loosely coupled solution works for non-rigidly attached sensors, which is more appropriate for pedestrians [156–159]. For humans, each part of the body goes through different motion dynamics, requiring the IMU and the camera to be separated for accurate position and orientation estimates at their appropriate mounting locations. To achieve higher accuracy with an IMU from human motion, the optimal mounting location is the foot, as this enables accurate detection of a motion constraint (stance phase) [12, 15, 160]. However, for a camera, a mounting location on the body other than the head could result in a blocked field of view by other moving limbs, or the camera could hinder the movement, for example if placed on the foot [161, 162]. Therefore, we have chosen to separate the camera and IMU to capture the motion dynamics accurately and still have a clear view of the environment. To combine the camera and IMU, a novel sensor fusion algorithm was devised using the motion constraint (stance phase) to loosely couple the visual and inertial sensors with an extended Kalman filter.

This chapter focuses on the implementation details of the novel indoor positioning system, and its evaluation. The chapter begins by describing the implementation details of the two components of the sensor fusion system: localisation with a foot-mounted IMU using the zero-velocity update (ZUPT) method (Section 4.1), and a camera pose estimation from fiducial markers (Section 4.2). Afterwards, the chapter explains the loosely coupled sensor fusion approach (Section 4.3). This is followed by the experimental setup, system design, ground truth method (Section 4.4), and key performance metrics (Section 4.5). Subsequently, an evaluation study (Section 4.6) and conclusion (Section 4.7) are presented.

4.1 Localisation with the Foot-Mounted IMU

During walking, the human feet alternate between swing (moving) and stance (stationary) phases in a periodic manner. Localisation with a foot-mounted IMU involves accurately detecting the stance phase and applying the ZUPT algorithm to correct the accumulated error since the last stance phase. The stance phase detection of our positioning system is based on Skog et al.'s work [13], which uses three inputs: acceleration magnitude, acceleration variance, and angular velocity. Figure 4.1 illustrates our implementation of the ZUPT-based localisation approach. The algorithm was implemented with an extended Kalman filter (EKF), and the error in position (p^n), velocity (v^n), and orientation matrix (C_b^n) values were fed back into the EKF to correct their values. The n denotes the navigation frame of reference, and b signifies the body frame of reference. Navigation and body frame of reference were explained in Section 3.1.

Orientation matrix (C_b^n) was the direction cosine matrix (DCM), and it is the product of roll (ϕ), pitch (θ), and yaw (ψ) rotation matrices which are IMU orientations in x, y, and z axes, respectively. The product of the three rotation matrices is shown in:

$$C_n^b = C_\psi C_\theta C_\phi = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\phi & sin\phi \\ 0 & -sin\phi & cos\phi \end{bmatrix} \begin{bmatrix} cos\theta & 0 & -sin\theta \\ 0 & 1 & 0 \\ sin\theta & 0 & cos\theta \end{bmatrix} \begin{bmatrix} cos\psi & sin\psi & 0 \\ -sin\psi & cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.1)$$

, and the DCM is given by [163]:

$$C_b^n = \left[C_n^b\right]^T = \begin{bmatrix} cos\theta cos\psi & sin\phi sin\theta cos\psi - cos\phi sin\psi & cos\phi sin\theta cos\psi + sin\phi sin\psi \\ cos\theta sin\psi & sin\phi sin\theta sin\psi + cos\phi sin\psi & cos\phi sin\theta cos\psi - sin\phi sin\psi \\ -sin\theta & sin\phi cos\theta & cos\phi cos\theta \end{bmatrix} \tag{4.2}$$

The initial value of the DCM was calculated from the accelerometer data when the IMU was in a stationary state. Subsequently, DCM is propagated using the gyroscope's angular velocity (ω^b) with:

$$C_b^n(k+1) = C_b^n(k) * [(\omega^b(k+1)\times) * dt_k] \tag{4.3}$$

using

$$[\omega^b(k+1)\times] = \begin{bmatrix} 0 & -\omega_z^b(k+1) & \omega_y^b(k+1) \\ \omega_z^b(k+1) & 0 & -\omega_x^b(k+1) \\ -\omega_y^b(k+1) & \omega_x^b(k+1) & 0 \end{bmatrix} \tag{4.4}$$

The step length, s_k, and step heading, ψ_k are calculated from one stance phase to the other. The stepping frequency and the IMU's update rate determine the next localisation state values using Equation 4.3 and

$$p^n(k+1) = p^n(k) + \frac{(v^n(k) + v^n(k+1))}{2} * dt_k, \tag{4.5}$$

$$v^n(k+1) = v^n(k) + (C_b^n(k) * a^b(k+1) - g) * dt_k) \tag{4.6}$$

where k is the sample index, dt_k is the time difference since the last update, $a^b(a_x^b, a_y^b, a_z^b)$ is the accelerometer's acceleration measured in the body frame of reference, $\omega^b(\omega_x^b, \omega_y^b, \omega_z^b)$ is the gyroscope's angular velocity measured in the body frame of reference and $g^b(g_x^b, g_y^b, g_z^b)$ is acceleration due to the gravity. The foot orientation values at a sample index k are given by:

$$Roll, \phi_k = arctan(\frac{C_{32}}{C_{33}}), \tag{4.7}$$

$$Pitch, \theta_k = -arcsin(C_{31}), \tag{4.8}$$

and

$$Yaw, \psi_k = arctan(\frac{C_{21}}{C_{11}}) \tag{4.9}$$

where C_{rc} refers to element of the DCM (Equation 4.2) at row (r) and column (c).

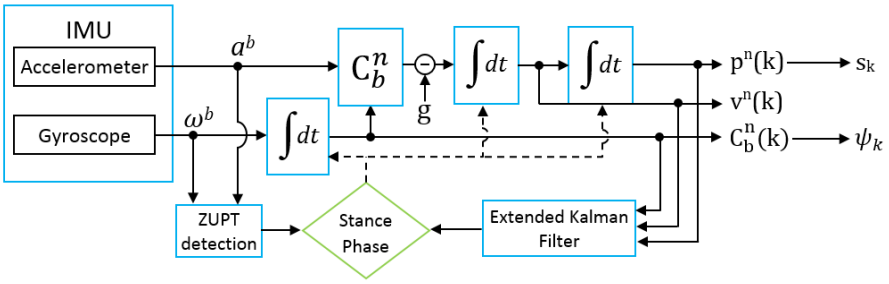


FIGURE 4.1: Block diagram of the ZUPT EKF correction mechanism

The IMU used for our system was XSens MTW Awinda [29]. Initially, TDK, Bosch, and STMicroelectronics IMUs were also considered, however, past research by Wagastaff et al. [14] and IMU specifications [26–29] indicated higher noise in accelerometer and gyroscope signals; therefore, we chose XSens MTw Awinda IMU. Figure 4.2 shows how the IMU was strapped near the toes of the pedestrian's foot. To decide on the sensor placement, we consulted Wahlström, et al.'s work which showed placing IMU near the toes or heel led to most accurate localisation results compared to the ankle and shoelaces location [164]. For our system, toes were selected as it is easier to mount the IMU near the toes than the heel. Another important aspect of the sensor placement is that the IMU should be tightly attached to the foot for accurate stance phase detection. Before each repetition of the experiment, we checked the Velcro strap holding the IMU to ensure it was tightly held. For consistency in between the experiments, the IMU was always mounted on the right foot of the subject with the LED facing forward - this kept the y-axis of the IMU oriented in the walking direction.

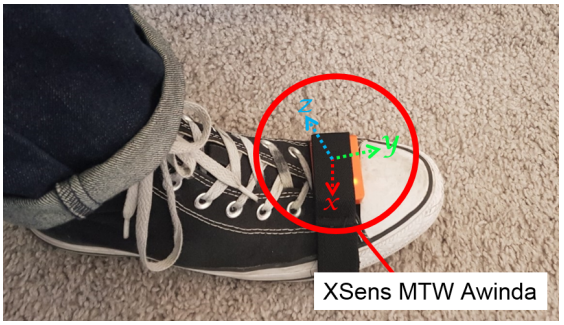


FIGURE 4.2: XSens MTW Awinda IMU on the right foot of a participant, and the IMU frame of reference shown by red(x-axis), green(y-axis), and blue(z-axis) arrows

4.2 Camera Pose Estimation from Fiducial Markers

The camera pose (position and orientation) estimation for our positioning system involved the following steps: detecting installed fiducial markers, computing relative pose of the camera from the markers, and calculating absolute camera pose. The camera pose transformation process is illustrated in Figure 4.3 and is detailed further in Section 4.2.2. To estimate the camera pose, fiducial markers were installed on the ceiling of the building. Each marker had a unique pattern and an identification number that was identifiable through binary image analysis on the captured video frames. All the markers also had a known absolute position and orientation value in the World Coordinate System (WCS), which was recorded during the offline mapping process (explained in Section 4.2.1). The WCS is an arbitrary reference system that starts from an assigned origin in a physical space. Each marker's absolute position and orientation was defined relative to the origin and was given by the (T_w^m) matrix. The T_w^m was used to transform the camera's relative pose, T_m^c, to an absolute pose value, T_w^c using:

$$T_w^c = T_w^m * T_m^c \tag{4.10}$$

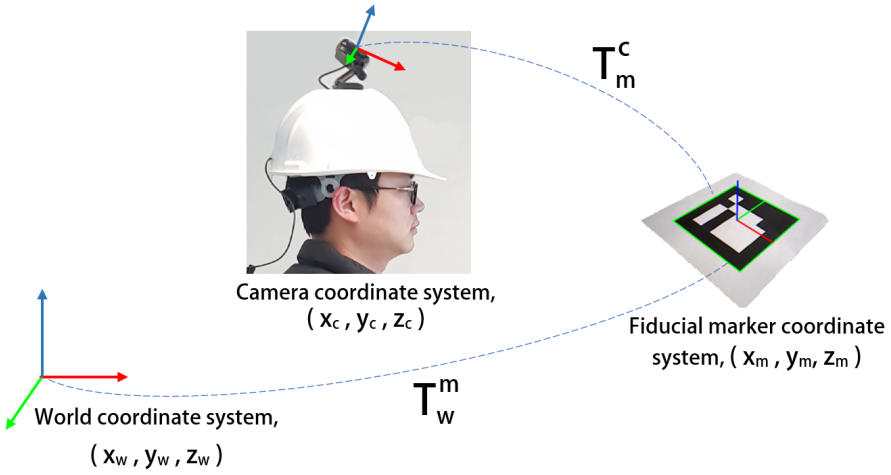


FIGURE 4.3: Camera pose transform from fiducial marker to WCS

4.2.1 Fiducial Marker Mapping

Accurate camera pose estimation requires an accurate map of the fiducial marker pose in the WCS. Marker mapping was performed using the setup shown in Figure 4.4. For each fiducial marker, four points (three corners and one centre) were marked from the ceiling to the ground using a Bosch three point laser mounted on a surveyor tripod. The

centre point of the marker was used to measure its position from the origin of the WCS with the Bosch PLR 50C laser measure. Orientation was calculated from three corner points. Each marker's transform (rotation and translation) was recorded in a mapping file with the marker identification number to be used later for localisation. Marker pose and data stored in the mapping file are shown in Figure 4.5. For our fiducial marker system, OpenCV's ArUco fiducial marker library was chosen due to its robustness to lighting conditions and easier scalability if more markers were needed [165, 166]. The ArUco marker dictionary used for our system was "DICT_7X7_100" [167].

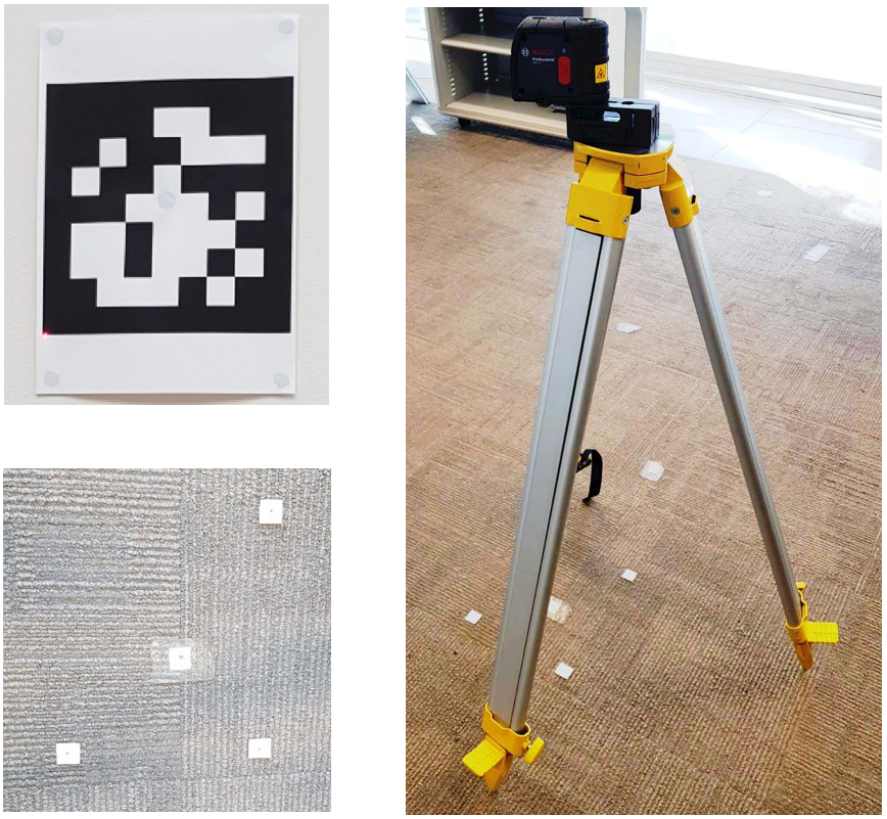


FIGURE 4.4: Mapping setup: surveyor tripod with mounted Bosch three point laser. The fiducial marker on the left has a red laser dot and corresponding markers on the ground for three corner points to measure the position and orientation

```
# MARKERS_MAPPING_FILE.TXT

75
1.00   0.00   0.00    0.72
0.00   1.00   0.00   12.44
0.00   0.00   1.00    5.15

74
1.00   0.00   0.00    0.72
0.00   1.00   0.00   11.92
0.00   0.00   1.00    5.15

8
1.00   0.00   0.00    0.41
0.00   1.00   0.00   11.47
0.00   0.00   1.00    5.15
```

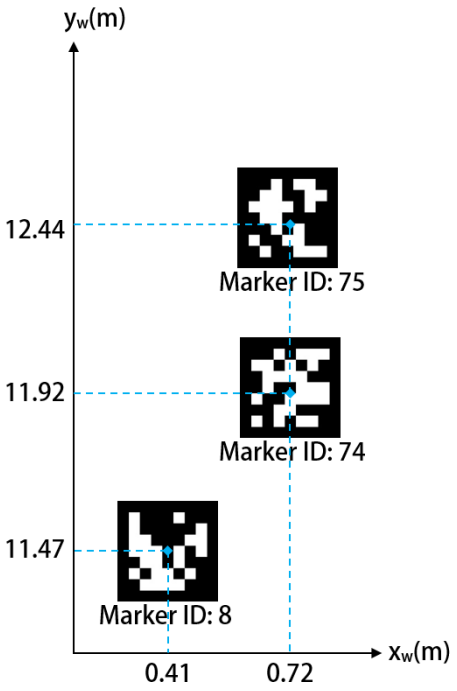


FIGURE 4.5: Marker mapping file containing the marker id in the first row and measured position and orientation values in subsequent rows

4.2.2 Estimating Camera Pose from the Detected Markers

After a marker was detected within the camera's video frame, its four corner points were used to compute the camera's relative pose, and the identification number was used to find the absolute transformation matrix T_w^m (marker to wcs) in the mapping file. Figure 4.6 shows the 3D transformation of fiducial marker's four corner points to a video frame. The camera's relative pose T_m^c was derived from a pinhole camera model [144], and is given by:

$$s * \begin{bmatrix} u_{ci} \\ v_{ci} \\ 1 \end{bmatrix} = K * T_m^c * \begin{bmatrix} cx_i \\ cy_i \\ cz_i \\ 1 \end{bmatrix} \tag{4.11}$$

where s is the projection scaling factor calculated during the camera calibration process described in Section 3.3, and u_{ci}, v_{ci} $(i = 1, ..., 4)$ are the four corners of one marker in the image. The cx_i, cy_i, cz_i $(i = 1, ..., 4)$ are the 3D coordinates of the marker in the fiducial coordinate system (fcs), and K is the camera's intrinsic matrix also calculated during the camera calibration process. The flow chart in Figure 4.7 illustrates the complete process of camera pose estimation from fiducial markers. The camera's absolute pose,

(T_w^c), can be determined with camera's relative pose T_m^c and marker's absolute pose T_w^m using Equation 4.10. To increase the accuracy of the localisation system, a higher density of fiducial markers were installed on the ceiling, which allowed multiple marker detection in one video frame.

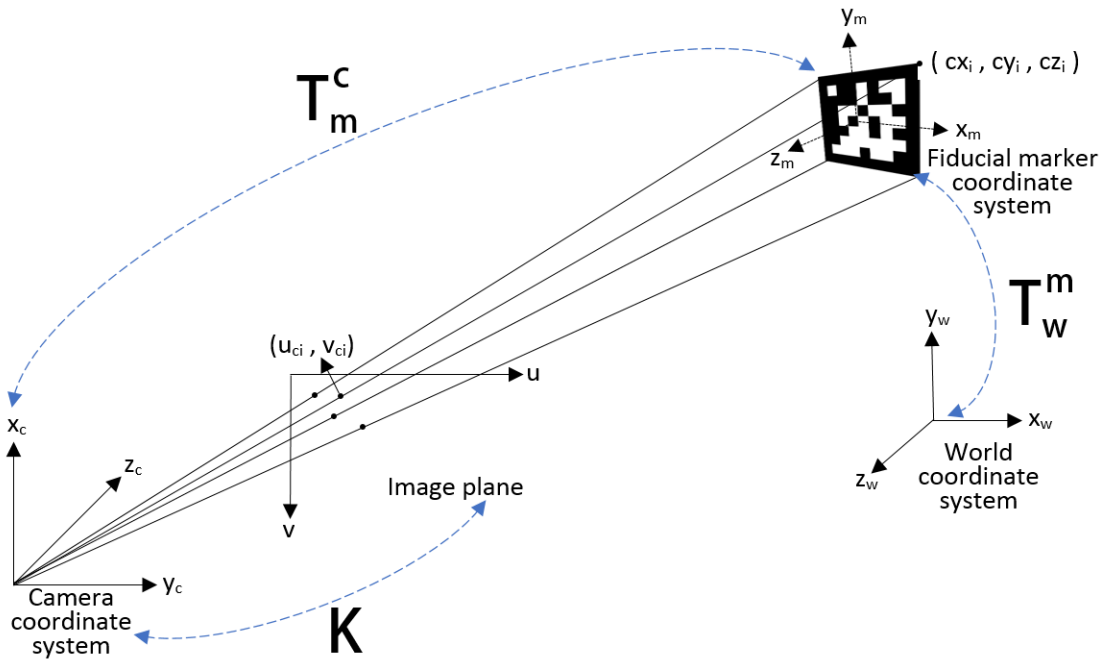


FIGURE 4.6: Fiducial marker corner points transformation in a pinhole camera model

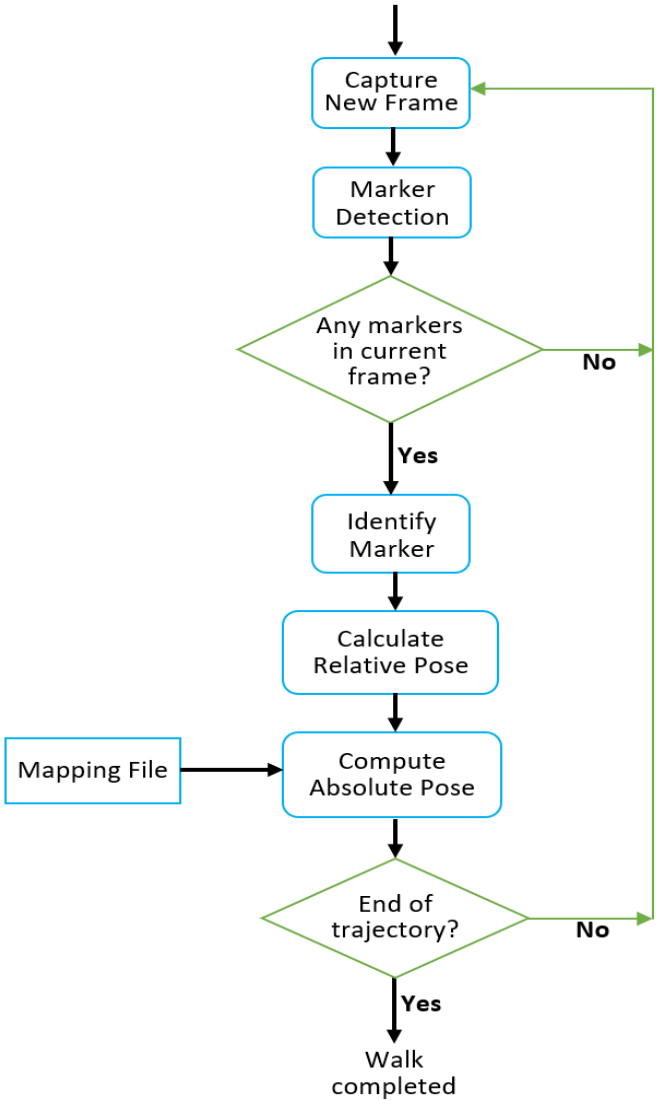


FIGURE 4.7: Flowchart of the camera pose estimation process

4.3 Loosely Coupled Sensor Fusion Method

Even after applying error correction during ZUPT, the localisation accuracy of the foot-mounted IMU deteriorates. To check the error growth, we proposed a sensor fusion approach between the foot-mounted IMU and head-mounted camera. The correction from the camera was applied during the stance phase when both foot and head were vertically aligned [168–171]. Both sensors were loosely coupled using an EKF. The flowchart in Figure 4.8 illustrates the implementation of our system. The motion model of the EKF was driven by the foot-mounted IMU data, and the measurement model by the camera-based localisation system. The state vector x_k of the localisation system

consisted of three components. The first two contained the 2D position px_k, py_k of the person, and the third contained the heading ψ_k value. The vector is described by:

$$x_k = (px_k, py_k, \psi_k)^T \tag{4.12}$$

where T is the transpose of the matrix, and k is the sample index, corresponding to time t_k. The motion model (which used the foot-mounted IMU data) was propagated with the function, $x_k = g(x_{k-1}, u_k, w_k)$, which is given by:

$$x_k = x_{k-1} + (s_k * cos(\psi_k), s_k * sin(\psi_k), \psi_k)^T + w_k \tag{4.13}$$

where w_k is the motion model noise. The s_k was the step length and was computed using the ZUPT approach described in Section 4.1. The covariance matrix P_k was propagated using the Jacobian matrix G_k and the motion model noise matrix, $Q = E[w_k * w_k^T]$. The G_k is given by:

$$G_k = \begin{pmatrix} 1 & 0 & -s_k * sin(\psi_k) \\ 0 & 0 & s_k * cos(\psi_k) \\ 0 & 0 & 1 \end{pmatrix}, \tag{4.14}$$

and Q is computed with the accelerometer and gyroscope noise values. The equation for covariance matrix propagation is:

$$P_k = G_k * P_{k-1} * G_k^T + Q_k \tag{4.15}$$

The measurement model was updated by the camera-based localisation system and was propagated with the function, $z_k = h(x_k, v_k)$, where v_k is measurement noise, which is used to calculate the measurement noise covariance matrix, $R_k = E[v_k * v_k^T]$. The state and covariance matrices in the measurement model were propagated using:

$$x_{k+1} = x_k + K_k * (z_{k+1} - h(x_k)), \tag{4.16}$$

$$P_{k+1} = (I - K_k * H_k)P_k \tag{4.17}$$

where K_k is the Kalman gain and can be calculated by:

$$K_k = P_k * H_k^T (H_k * P_k * H_k + R_k)^{-1} \tag{4.18}$$

The H_k is the Jacobian matrix of the function $h(x_k)$, and it was calculated by the partial derivative of $h(x_k)$ with respect to state x. It was evaluated using the prior state estimate x_{k-1}, and is given by Equation 4.18 and

$$H_k = \left.\frac{\partial h}{\partial x}\right|_{x=x_{k-1}} \tag{4.19}$$

Before localising with the sensor fusion system, a temporal offset between the camera and IMU was calculated offline, the IMU's accelerometer and gyroscope biases were removed, and the initial position and orientation of the localisation system were determined using initially detected fiducial markers. Even though we used one system clock (laptop's system clock) to record the camera's and IMU's timestamps, the recorded timestamps were not equal to the time instant when the sensors' data were sampled. Due to the transmission and triggering delay and unsynchronised clock at the camera and IMU end, there is a fixed temporal offset between the sensors that needed to be calculated and accounted for in the localisation system. We used Kalibr toolbox [172, 173] to determine the temporal offset and calibrated our system with the calculated value. Calculation of IMU's biases and initialisation of the localisation system were other calibration steps performed for our system. Before each experiment, the tracked pedestrian waited for 20.0 seconds in a stationary state, and the recorded acceleration and angular velocity data during the 20.0 seconds were used to calibrate for the biases. The camera pose estimated from the detected fiducial markers during this time was used to initialise the localisation system. After these calibration steps, the foot-mounted IMU calculated the change in step length and heading values using the IMU data. On stance phase detection, ZUPT was applied and the camera system was checked for any detected markers. If there were detected markers, camera pose was estimated using the method described in Section4.2.2. EKF's measurement model was applied using the computed camera pose value. The estimated camera pose was used to calculate the position and heading values of the localisation system. The process continued to the end of the trajectory.

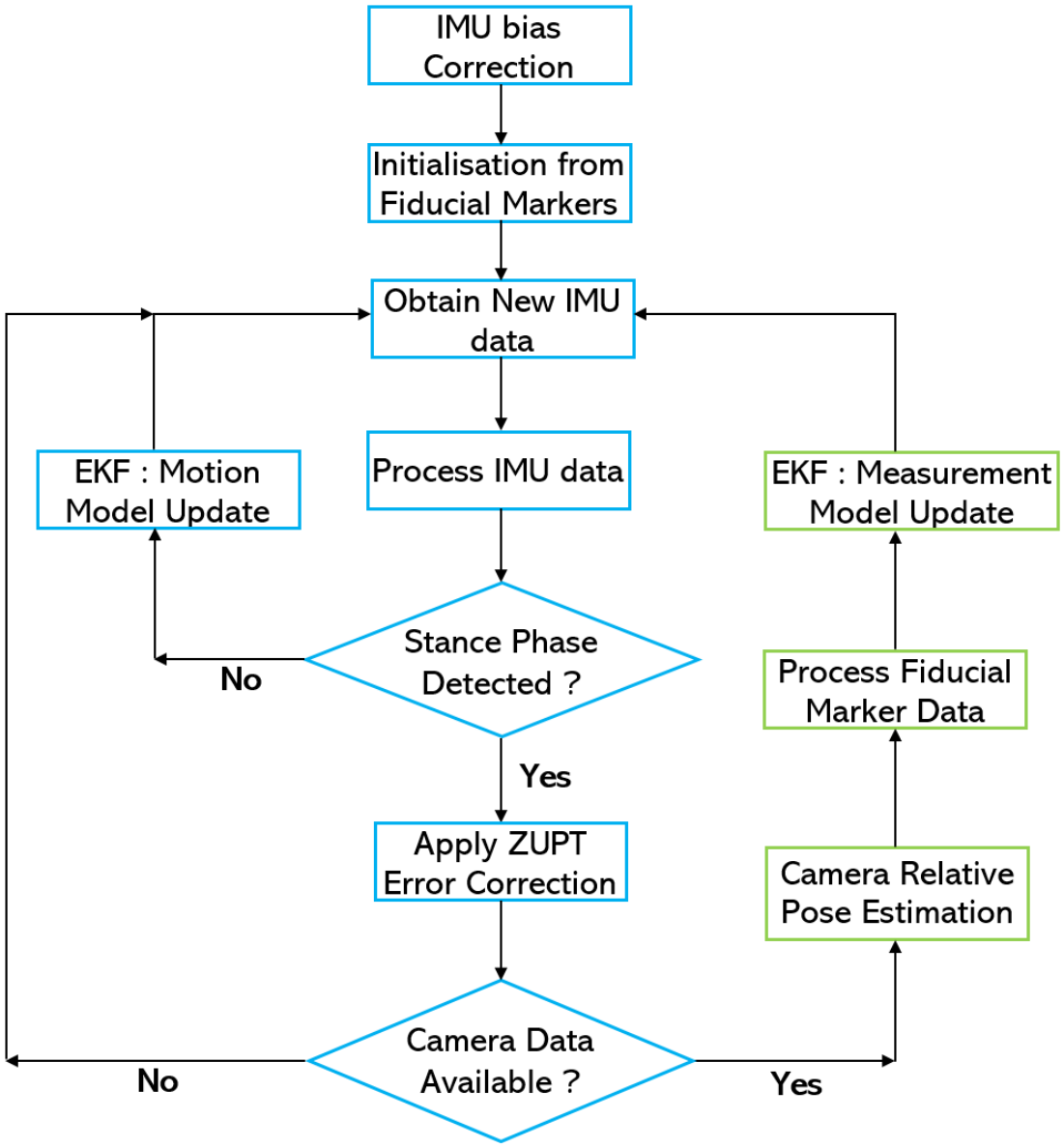


FIGURE 4.8: Sensor fusion method flow chart

4.4 Experimental Setup

An indoor area of dimension $17m$ x $9m$ was chosen inside the HIT Lab NZ building to install the markers for large-scale localisation. The floor plan of the tracking area and the planned walking trajectory are shown in Figure 4.9. In total 66 markers were installed with an average spacing of $0.5m$. Figure 4.10 shows the marker installation in the tracking space.

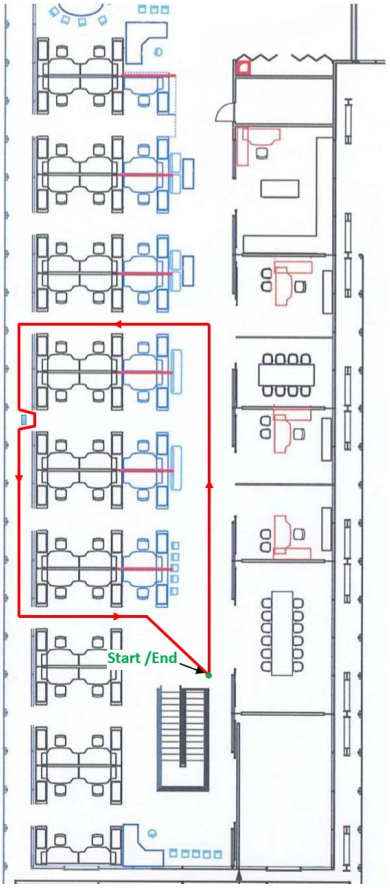


FIGURE 4.9: Indoor tracking area and the planned walking trajectory

FIGURE 4.10: Installed markers in the tracking space

4.4.1 Hardware

The hardware components of the developed system consisted of three devices: an XSens MTW Awinda IMU, a Logitech C920 camera, and a Razer blade 13 *inch* laptop. The specifications of the three devices are in Table 4.1, Table 4.2, and Table 4.3 respectively.

TABLE 4.1: XSens MTW Awinda IMU specifications

Specifications	**Accelerometer**	**Gyroscope**	**Magnetometer**
Size	4.7 x 3.0 x 1.3 cm		
Weight	16 g		
Sampling frequency	20-120 Hz	20-120 Hz	20-120 Hz
Operation temperature	0-50 ^{o}C	0-50 ^{o}C	0-50 ^{o}C
Full scale (FS)	$\pm$160 m/s^2	$\pm$2000 deg/s	$\pm$1.9 $Gauss$
Non-linearity	0.5% of FS	0.1% of FS	0.1% of FS
Bias stability	0.1 mg	10 $deg/hour$	-
Noise	200 $\mu g/\sqrt{Hz}$	0.01 $degs^{-1}/\sqrt{Hz}$	0.2 $mGauss/\sqrt{Hz}$

TABLE 4.2: Logitech C920 camera specifications

Specifications	
Size	4.3 x 9.4 x 7.1 cm
Weight	162 g
Resolution	1920x1080
Frame Rate	30 Hz
Field of View	78^o

TABLE 4.3: Razer blade 13 inch laptop specifications

Specifications	
Operating System	Windows 10 (64-bit)
CPU	Core i7-8565U, 1.8 GHz
RAM	8 GB
Hard Drive	256 GB
GPU	Intel UHD Graphics 620
Size	30.5 x 21.0 x 1.5 cm
Weight	1280 g

4.4.2 Ground Truth Methods

To choose the ground truth method, we analysed past research which revealed four main methods being used. The first method is infrared markers-based motion capture system, such as Vicon or OptiTrack [174, 175]. These systems work well in environments that have limited exposure to sunlight, as this interferes with the infrared communication between the markers and camera. The tracking space for our system is illuminated by sunlight making this method unviable for our scenario. Additionally, the cost of these systems is high. To cover a $15m$ x $15m$ tracking area, 20 OptiTrack Prime^x 22 or 12 OptiTrack Prime^x 41 cameras are required, which cost nearly $80,000 USD [176]. The second method has been used by Dina et al., and involves accurately installing markers on the ground and using a button press on passing the markers [160]. As this requires input from the user walking over the marker or a person following the user, it is prone to human reaction error at each marker point. The third method, used by Ryan et al., involves a subject followed by a person with a measuring wheel recording the distance travelled [177]. The main drawback of this method is that turn events are not registered, so it is only applicable for comparing total distances travelled. The fourth method has been used by Fusco et al., and it is similar to the second method described above. This method involves accurately placing markers on the floor and recording the tracked subject with an external camera [30, 178]. A video records the position and orientation of the subject with respect to the surrounding environment and the markers on the floor. The recorded video is analysed offline to compute the Euclidean distance error between the ground truth and estimated position by the designed system. This method requires a lot of time to analyse the video afterwards; however, it is less prone to estimation error.

Ground truth for our system was performed using the fourth method. Reference markers were installed on the floor at 30 cm intervals and their position was measured using a Bosch PLR 50C laser measure. For each experiment, tracked subject motion was recorded with an external GoPro Hero5 camera. The recorded video contained the position of the subject's feet in reference to floor markers and surrounding environment, and the screen of the laptop showing the ceiling fiducial marker video. The video was later analysed to record the ground truth data and compute the Euclidean distance error.

4.4.3 Data Collection and Processing

During the experiment, the subject wore the XSens MTW Awinda IMU on the right foot and the helmet with Logitech C920 camera facing the ceiling of the building. Both sensors recorded data along with the timestamp values on the laptop (Razer Blade 13.3 inch) carried by the subject. Figure 4.11 shows the system setup for data collection. The IMU was wirelessly connected to the laptop, and the camera was connected via a USB. Each sensor was operating at a different frequency; the IMU captured data at 100 Hz whereas the camera recorded video at 30 Hz with 1080p resolution. Both sensors were temporally aligned using the time offset calculated with the Kalibr toolbox. To remove the accelerometer and gyroscope bias from the IMU data and initialise the localisation system, an initial 20s of data was recorded with the subject in a stationary state. The bias removal method and initialisation were performed before each walk. The data from the IMU contained three-dimensional acceleration, three-dimensional angular velocity, and timestamp values. The data from the head-mounted camera video contained the fiducial marker ID along with the camera pose and timestamp values. The collected data was processed on MATLAB where it was combined using EKF.

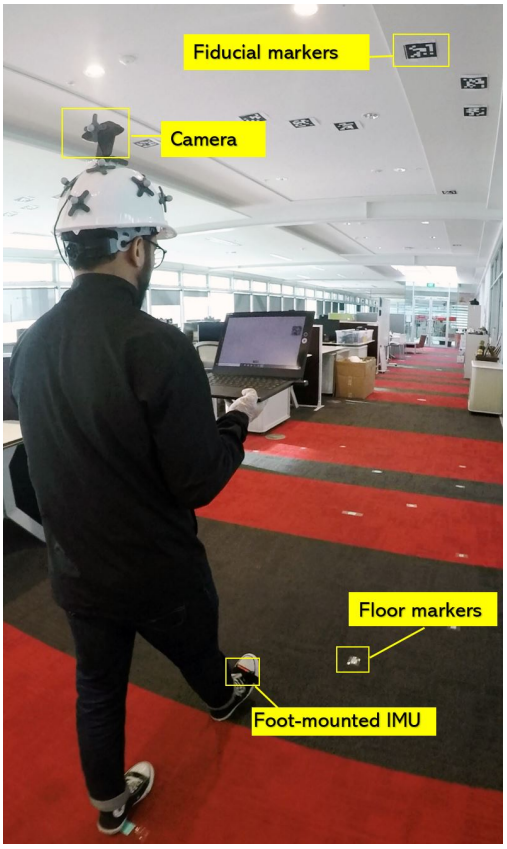


FIGURE 4.11: System setup

4.5 Key Performance Metrics

The main performance metrics used in this thesis to evaluate the localisation system are: positioning accuracy, precision, and energy efficiency. Details on these parameters are as follows.

4.5.1 Accuracy

In the indoor localisation literature, positioning accuracy has been measured in three different ways: position and heading sample points on different parts of the trajectory [14, 160]; the total distance travelled [179]; and the difference between the start and end point when starting and ending at the same location (also known as return position error) [180]. For each of these three methods, the positioning accuracy is calculated by measuring the localisation error e_i which is the Euclidean distance between the estimated position by the localisation algorithm $p_i = (x_i, y_i)$, and the ground truth $p_i^g = (x_i^g, y_i^g)$, and is given by:

$$e_i = \sqrt{(x_i^g - x_i)^2 + (y_i^g - y_i)^2} \tag{4.20}$$

where i is the sample index, ranging from 1 to m.

4.5.2 Precision

Precision value reflects the distribution of localisation error and reveals the overall performance of the system. The precision value in the indoor localisation literature has been mainly visualised using the empirical cumulative distribution function (eCDF) plots of the localisation error [164, 181]. The eCDF plot is a non-parametric estimation of the cumulative distribution function (CDF), and is calculated by arranging the localisation error e_i (Equation 4.20 from smallest to largest and assigning a probability of $1/m$ to each data point, where m is the total number of data points. This results in a step function which increases by $1/m$ at each data point. To compare the localisation systems, a $90th$ or 95th percentile is chosen on the eCDF plot and the corresponding localisation error is used to evaluate the system. For our system, we used the 95th percentile as a reference.

4.5.3 Energy Efficiency

The energy consumption of the system is estimated by measuring the power consumed by each component and multiplying it with the time taken. For our system, there are three main energy components: (i) energy consumed by the camera system to record and process the fiducial markers video, E_{cam}, (ii) energy required by the inertial sensors, E_{imu}, and (iii) the energy consumed by the sensor fusion algorithm, E_{fusion}. The total energy, E_{total}, is the summation of the three components and is calculated by:

$$E_{total} = E_{cam} + E_{imu} + E_{fusion} \tag{4.21}$$

Unlike accuracy and precision, energy efficiency is not used in this Chapter, as it is explained with other performance metrics. However, it is applied in Section 5.3 and Section 5.4.

4.6 Evaluation Study, Results, and Discussion

A study was conducted to evaluate the performance of the foot-mounted IMU, camera-based localisation system, and loosely coupled sensor fusion approach. For the study, we planned a $250m$ walking trajectory, consisting of five loops of a $50m$ path which had fiducial markers installed on the ceiling. The floor plan in Figure 4.9 shows the planned walking trajectory. Since the main goal of the study was long-term localisation accuracy, the subject walked only in a forward direction. There were three repetitions of the experiment, each consisting of a $250m$ forward walk. The data from the camera-based localisation system and foot-mounted IMU was recorded on the Razer blade laptop and later processed on MATLAB. Before each repetition, the subject was stationary for $20s$ to remove the IMU biases and record the initial orientation of the IMU for transformation between the body and navigation frames. The results of these experiments are presented in three subsections: localisation with only the foot-mounted IMU, Section 4.6.1; localisation with the camera-based localisation system, Section 4.6.2; and localisation with the loosely coupled sensor fusion approach Section 4.6.3.

4.6.1 Foot-mounted IMU Trajectories

Figures 4.12, 4.13, and 4.14 show the walking trajectories of the three repetitions of the experiment, which were estimated using the foot-mounted IMU data and by applying the ZUPT algorithm for correction. The blue line in the Figures represents the walked trajectory; the green line is the ground truth. The red and blue squares in the Figures are start and end positions, respectively. We also computed the localisation error for the three trajectories using the ground truth data and created a eCDF plot, which is shown in Figure 4.15. A fuller discussion of the foot-mounted IMU is in Section 4.6.4.

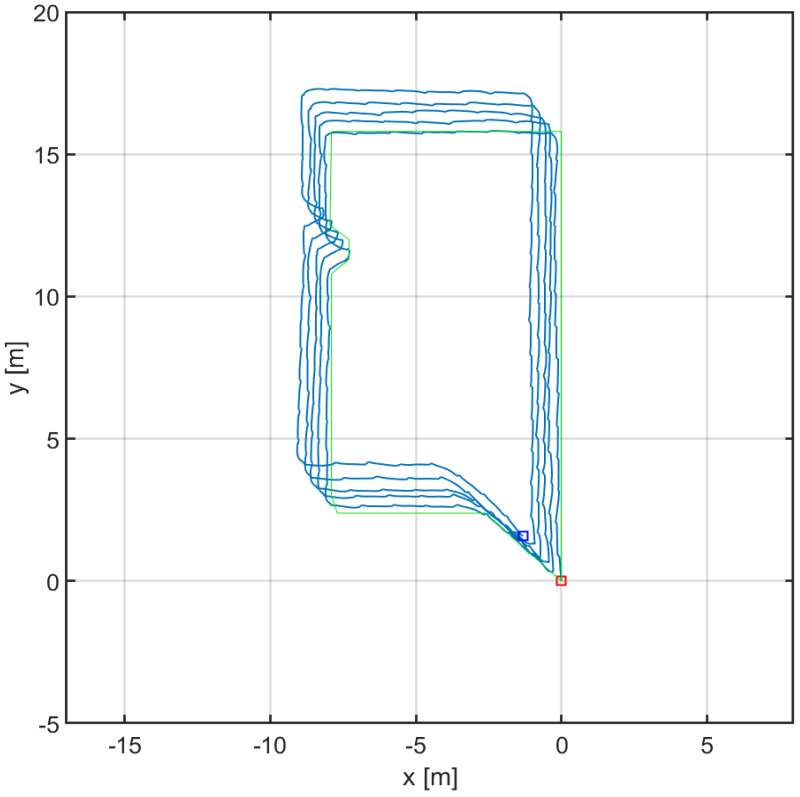


FIGURE 4.12: Log 1 of the foot-mounted IMU trajectory

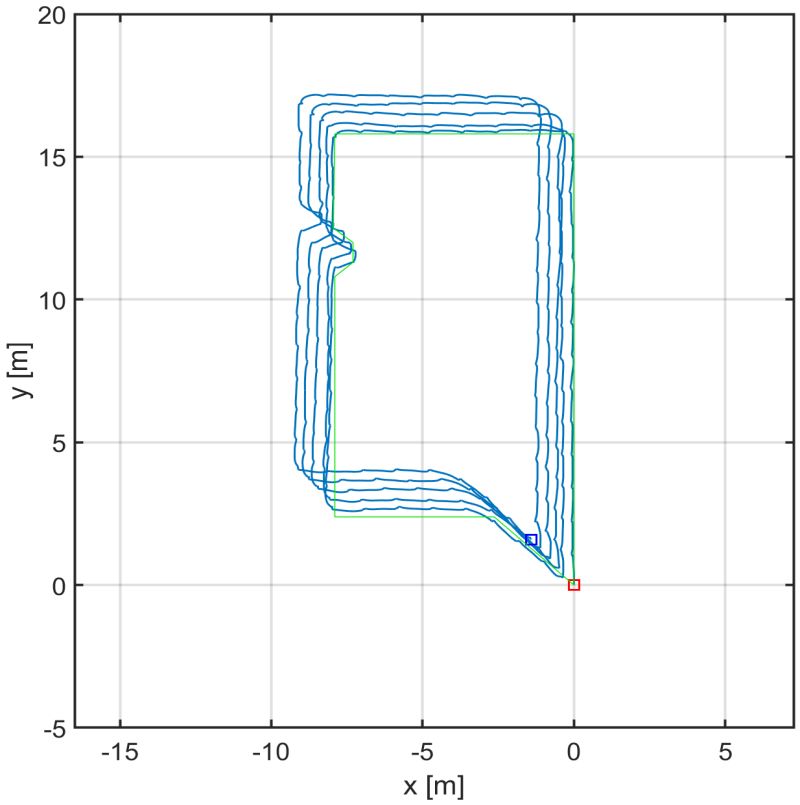


FIGURE 4.13: Log 2 of the foot-mounted IMU trajectory

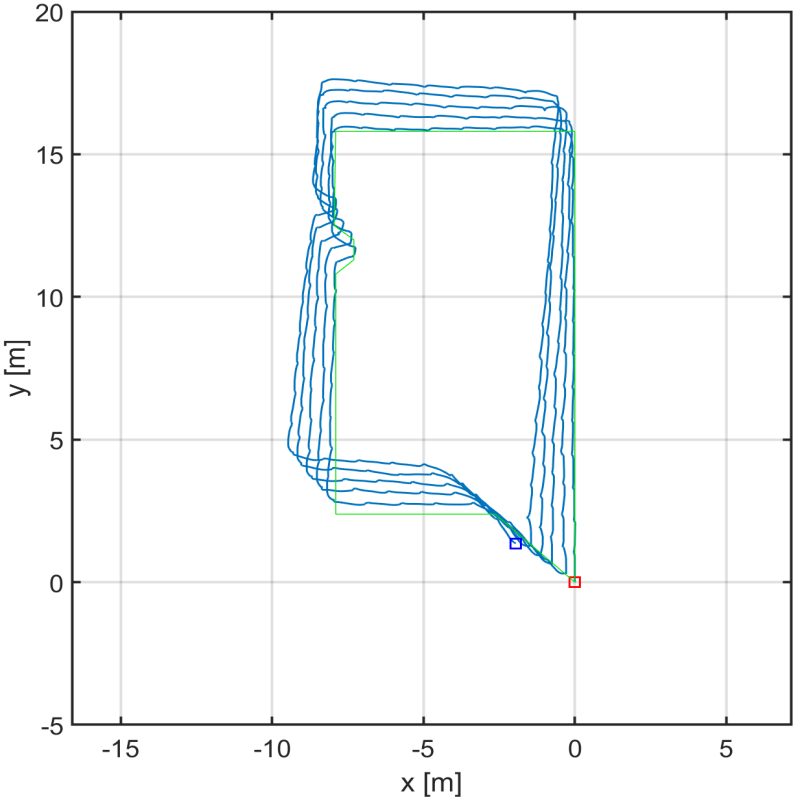


FIGURE 4.14: Log 3 of the foot-mounted IMU trajectory

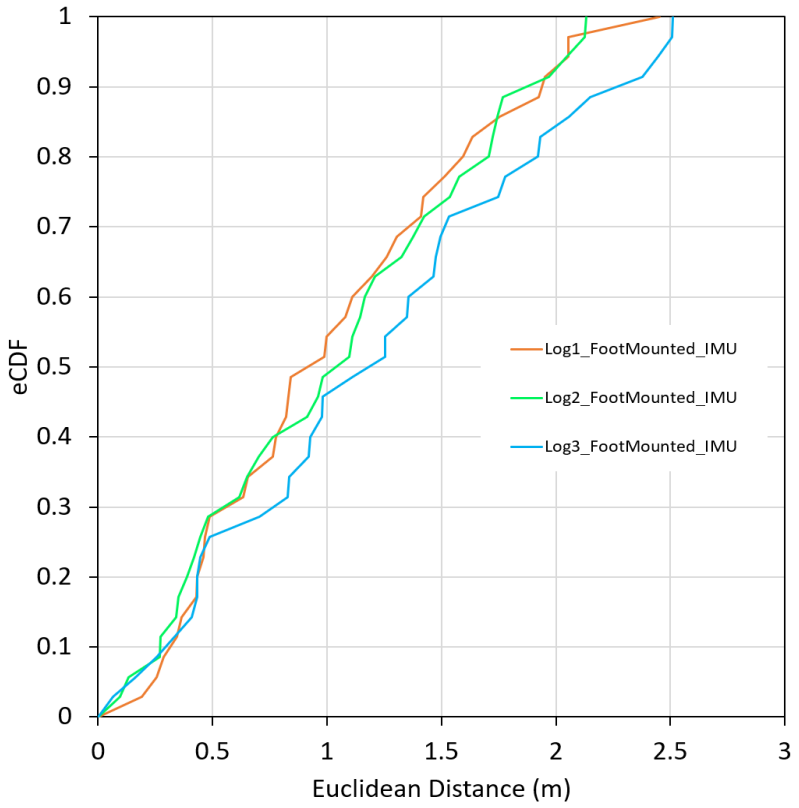


FIGURE 4.15: Localisation error eCDF plot of the foot-mounted IMU

4.6.2 Camera-based Localisation Trajectories

Figures 4.16, 4.17, and 4.18 show the walking trajectories calculated by the camera-based localisation system for the three repetitions of the experiment. Each marker is assigned a colour and is represented as a "*" symbol with its assigned colour. The subject's position calculated from the marker detection is shown by "." with the same colour as the marker. There is a number beside each marker indicating the marker identification number, and it is based on an ArUco library's "DICT_7X7_100" dictionary [167]. The ground truth trajectory is indicated by the green line. Similar to the foot-mounted IMU, we also computed the localisation error for the three trajectories, and the eCDF plot of the error is shown in Figure 4.19. A fuller discussion of the camera-based localisation system is in Section 4.6.4.

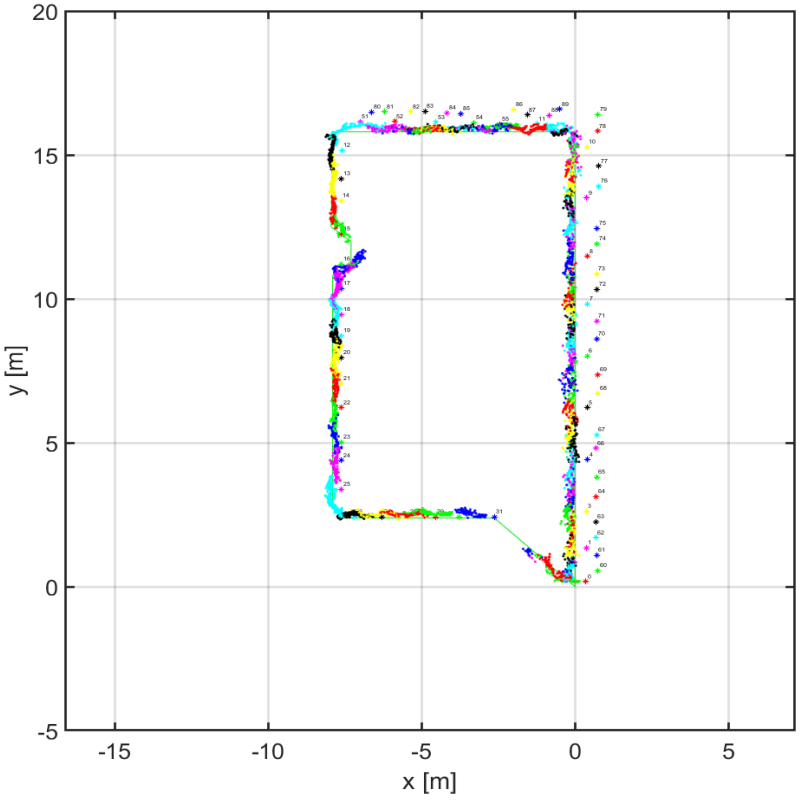


FIGURE 4.16: Log 1 of the camera-based localisation trajectory

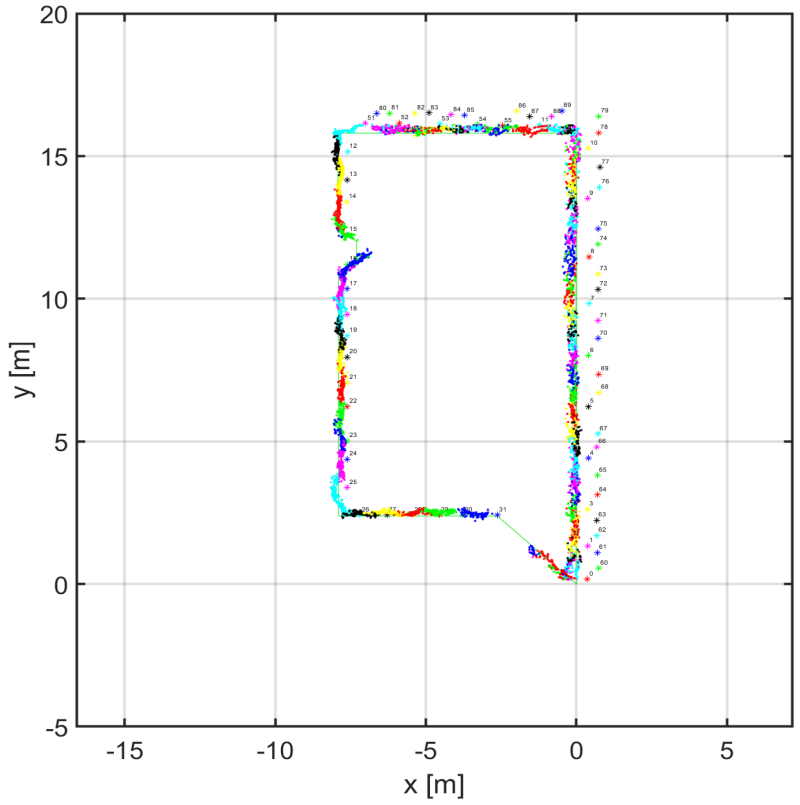


FIGURE 4.17: Log 2 of the camera-based localisation trajectory

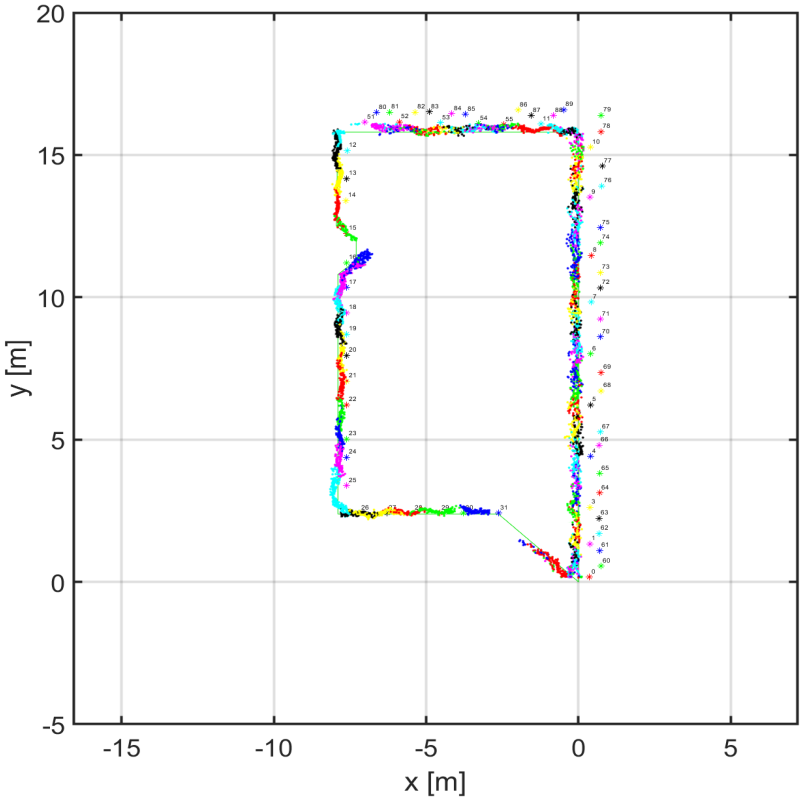


FIGURE 4.18: Log 3 of the camera-based localisation trajectory

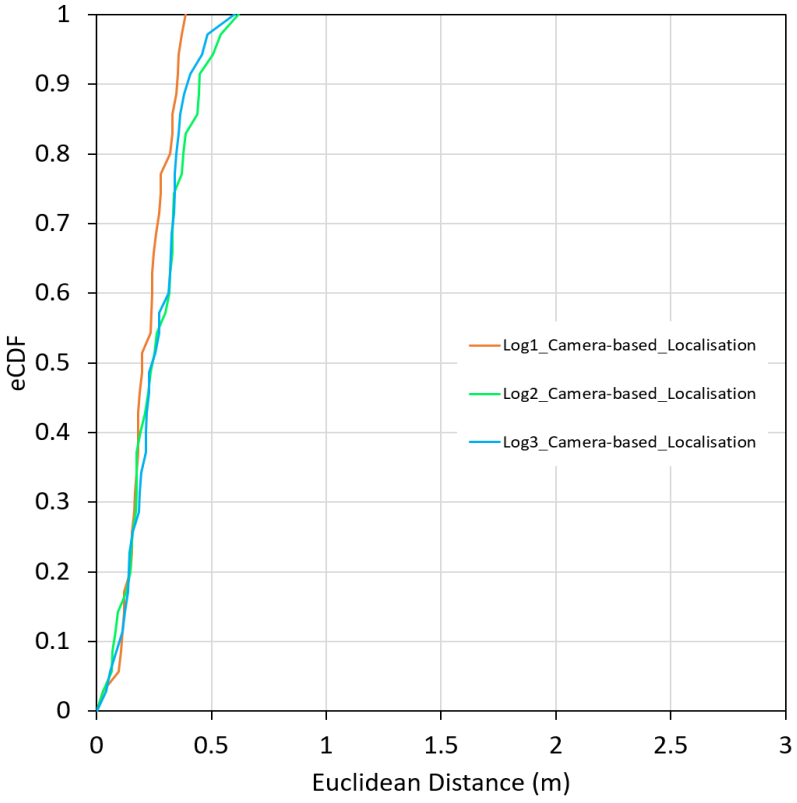


FIGURE 4.19: Localisation error eCDF plot of the camera-based system

4.6.3 Loosely Coupled Sensor Fusion Trajectories

Figures 4.20, 4.21, and 4.22 show the walking trajectories estimated by the sensor fusion algorithm for the three repetitions of the experiment. The purple line represents the walked trajectory, and the green line indicates the ground truth. The localisation error of the three trajectories is calculated using the ground truth data, and the resulting eCDF plot of the error is shown in Figure 4.23. A fuller discussion of the loosely coupled sensor fusion approach along with the foot-mounted IMU and camera-based localisation systems is in Section 4.6.4.

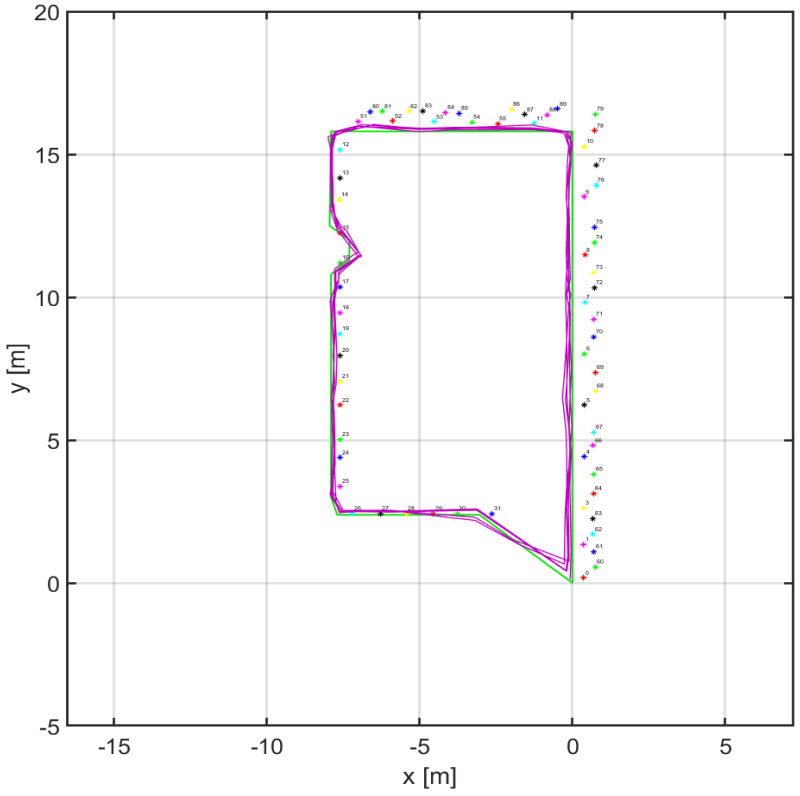


FIGURE 4.20: Log 1 of the loosely coupled sensor fusion trajectory

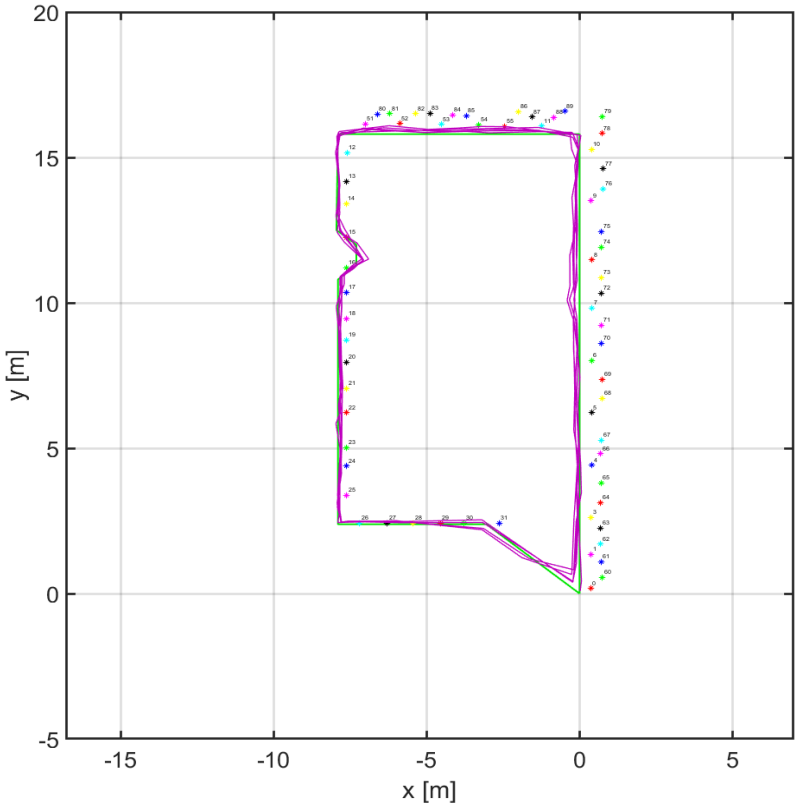


FIGURE 4.21: Log 2 of the loosely coupled sensor fusion trajectory

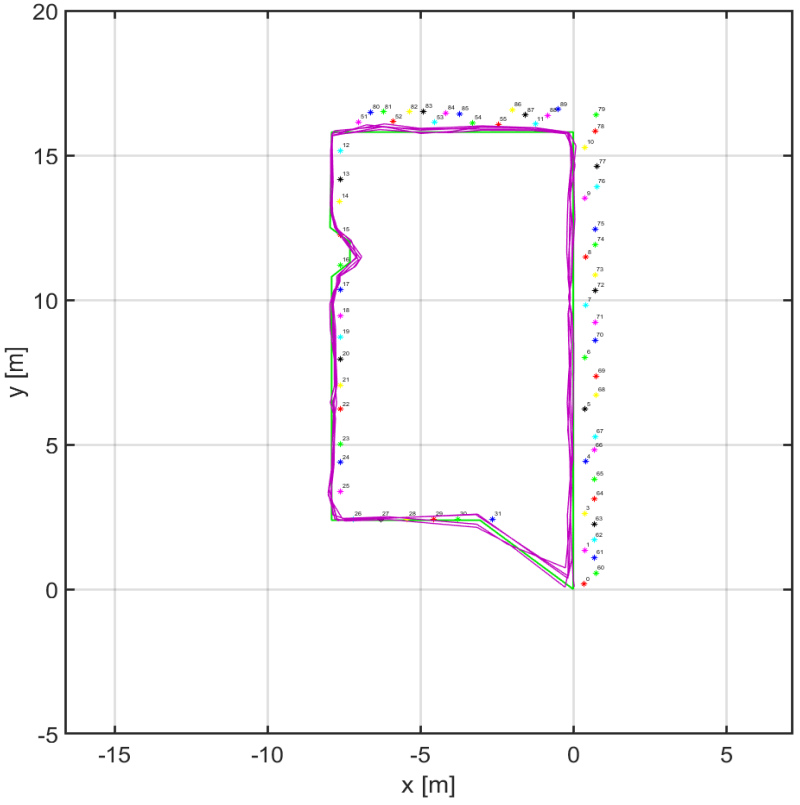


FIGURE 4.22: Log 3 of the loosely coupled sensor fusion trajectory

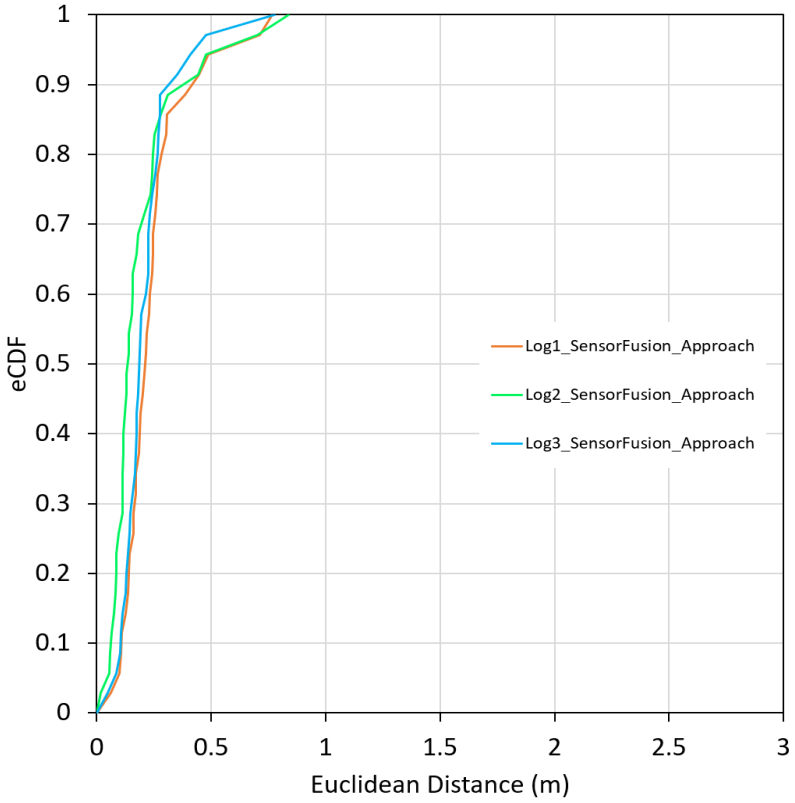


FIGURE 4.23: Localisation error eCDF plot of the sensor fusion method

4.6.4 Discussion

Results of the study indicate that the camera-based localisation provides accurate long term position tracking; however, it will lose tracking when markers are not in the camera's line of sight. As can be seen in the camera-based localisation trajectory plots, Figures 4.16, 4.17, and 4.18, the detected marker points closely follow the ground truth trajectory except in the areas where there are no markers detected, such as near the endpoint after marker 31 or between markers 15 and 16. With the loosely coupled sensor fusion approach, tracking in these parts is covered by the foot-mounted IMU as seen in Figures 4.20, 4.21, and 4.22. Localisation errors for Log 1, Log 2, and Log 3 of the camera-based localisation system at 95% of the walk are $0.35m$, $0.5m$, and $0.45m$, respectively, and for the loosely coupled sensor fusion approach $0.43m$, $0.55m$, and $0.56m$, respectively. From the foot-mounted IMU trajectories in Figures 4.12, 4.13, and 4.14, a general trend was observed: the tracked position represented by the blue line drifted away from the ground truth after each loop, indicating the error in stride length estimation and its cumulative effect on the preceding values. Change in orientation was

tracked accurately for Logs 1 and 2 of the foot-mounted IMU trajectories, but for Log 3, a gradual drift in orientation was observed due to the random noise in the gyroscope data, which is detailed in Section 3.2.1. Localisation errors for Log 1, Log 2, and Log 3 of the foot-mounted IMU trajectories at 95% of the walk are $2.05m$, $2.12m$, and $2.45m$, respectively.

To understand our results in the context of recent research, we made a comparative analysis. Due to the variation in total distances travelled in the different projects, we calculated the error percentage of the travelled distance. Table 4.4 shows the comparison of our system with similar recent work; it includes the system description of the positioning system, the motion types detected by the system, the total distance travelled during the experiment and the error percentage. When we compare our results in Table 4.4 with similar approaches, we find that our system is comparatively accurate. However, when we calculate the error percentage of the forward walk with a foot-mounted IMU and no correction from the camera-based localisation system, which is the case for [14, 182–184], it provides a comparable metric with forward walk having 0.88% error. With the passive infrastructure correction from fiducial markers, the overall system accuracy became better than [14, 182, 184–186], or comparable to [183], but increasing the overall power consumption of the system which in turn implies shorter battery life. The hybrid foot-mounted IMU and UWB system performed to a similar degree when we look at error to distance covered percentage; however, the data provided are only for a short walk $26m$ long. A longer walk test will provide more substantial evidence. In addition, the UWB system requires an active power supply at the infrastructure, which makes it less applicable for training exercises happening at a remote location. The sensor fusion between a foot-mounted IMU and WiFi fingerprinting-based positioning system has relatively lower accuracy due to the variation of access points available in the environment, which is also indicated by Chen et al. [185]. The choice of positioning technology with passive infrastructure is limited; however, the positioning results of our system indicate that it can potentially be used for applications such as training exercises that require sub-metre localisation accuracy and lack or have limited power infrastructure.

TABLE 4.4: Comparative analysis of localisation error with recent work.

Authors	**Ref.**	**System Description**	**Motion Types**	**Total Dist.** (m)	**Error %**
Our System	NA	Foot-mounted IMU and camera-based localisation system	Forward Walking	$250m$	0.21%
Shi et al.	[182]	Foot-mounted IMU	Forward Walking and Running	$100m$, $400m$	1.06%
Qiu et al.	[183]	Foot-mounted IMU	Forward Walking and Climbing	$50m$, $425m$	0.35%
Wagstaff et al.	[14]	Foot-mounted IMU	Forward Walking and Running	$130m$	2.47%
Rantanen et al.	[184]	Foot-mounted IMU	Forward Walking, Running, and Climbing	$170m$	2.64%
Chen et al.	[186]	Foot-mounted IMU and UWB-based positioning system	Walking	26 m	0.62%
Chen et al.	[185]	Handheld IMU device and WiFi fingerprinting-based positioning system	Walking	$90m$, $100m$, and $260m$	1.89%

4.7 Conclusion

This chapter described the implementation and evaluation of the loosely coupled sensor fusion system including two components: foot-mounted IMU and camera-based localisation systems. For foot-mounted IMU, we used the ZUPT algorithm to utilise a motion constraint (stance phase) during walking to determine the position of the subject. For the camera-based localisation system, the four corner points of fiducial markers were used to estimate the relative camera position and the mapping file to calculate the absolute camera position. Both systems were loosely coupled with an extended Kalman filter to provide accurate long term tracking. At the end of the chapter, we evaluated the sensor fusion system and the individual systems for localisation accuracy. The loosely coupled sensor fusion and camera-based localisation trajectories were within $0.6m$ error for 95% of the walk, and the foot-mounted IMU trajectories were within $2.45m$ error for 95% of the walk.

We have demonstrated that long term sub-metre accuracy can be achieved with the loosely coupled foot-mounted IMU and camera-based localisation system. However, to run the camera and process the image data requires a lot of energy; additionally, to obtain accurate results fiducial markers have to be within the camera's field of view. A foot-mounted IMU with the ZUPT method can achieve below $3m$ localisation accuracy but not the required sub-metre level. However, the energy consumption of the foot-mounted IMU is 16 to 30 times less than the camera-based localisation system, and no external reference is needed to compute the position and orientation value. If the localisation from the foot-mounted IMU is discretely corrected by a camera-based localisation system which goes to sleep mode between the corrections, the sensor fusion system could achieve sub-metre accuracy along with lower energy consumption. This is explored further in Chapter 5.

Chapter 5

Energy Efficient Localisation using Distance-based Camera Correction

The previous chapter described the implementation of the loosely coupled sensor fusion approach. It also presented a study to evaluate the long-term positioning accuracy of the devised localisation system. The results showed a sub-metre accuracy for the loosely coupled and camera-based localisation systems, whereas localisation with the foot-mounted inertial sensor produced 2-3 metres of positioning error over a $250m$ walk. However, the energy consumption of the camera-based system is significantly higher than the foot-mounted IMU. Having higher energy consumption implies that user's battery will last for a shorter duration and requires frequent recharging, which becomes a limitation during training exercises, especially when there are many trainees involved. Unlike localisation accuracy, there has been limited research on improving energy efficiency of the positioning system (refer to Section 5.1 for more details). This chapter proposes a novel distance-based error correction approach which uses distance estimated by the foot-mounted inertial sensor to reduce the number of corrections required from the visual sensor. This chapter also explores the energy/accuracy trade-off by varying the distance of correction from the visual sensor, and observing its effects on energy consumption and overall positioning accuracy. More frequent corrections at a shorter distance result in a more accurate solution, but increase the energy consumption as well

which is not suitable for battery powered devices. Apart from the energy efficiency, this chapter also studies the long-term effect of different walking motion types (forward, backward, and sideways walk) on the positioning accuracy, and uses these motion types to evaluate the distance-based error correction approach.

This chapter begins with the background and related work on energy-efficient localisation systems (Section 5.1), which is followed by the description of the proposed distance-based error correction approach (Section 5.2). Subsequently, the method and instruments used to accurately measure the energy consumption of the system are explained (Section 5.3). Afterwards, the chapter describes the experiment conducted to evaluate the proposed method. Then the results, discussion (Section 5.4), and conclusion are presented (Section 5.5).

5.1 Background

Positioning accuracy has been extensively studied in the literature; however, there is limited research focused on energy efficiency of the localisation system. Existing work has been mainly done on radio-inertial [187–189], radio frequency only [190, 191], or inertial only [25, 192] localisation solutions. To the best of our knowledge, a visual-inertial localisation system has not been studied from the energy efficiency perspective. We will briefly go through the past research to understand different energy saving approaches used by these systems.

There are mainly two types of energy conservation methods found in the literature: (1) duty cycle driven methods, (2) data driven methods [193]. Duty cycle driven methods involve fixing the sample frequency of a device based on a user selected mode and keeping the sampling frequency constant during the tracking period. They are common in wearable devices, such as smartwatches or fitness wrist bands, which contain only inertial sensors for localisation [194]. For these devices, normally a user selects a motion type (such as running, walking, or swimming) which determines the constant sampling frequency during the localisation phase. The advantage of a duty cycle driven method is simpler implementation and no computation is required to vary the sampling frequency of the sensor or to make a decision to take an external input. Reducing the duty cycle

conserves the energy and the battery lasts longer; however, the disadvantage is the accuracy deteriorates especially for long-term tracking, which results in a large localisation error [195]. The second type, data driven methods, use a computed or external data input to control the duty cycle of the system or allow a correction from an external source. These are common in radio-inertial or radio only localisation solutions [188]. We also used a data driven approach where distance computed from the foot-mounted IMU decided the correction from the camera-based localisation system.

Another novelty explored in this chapter is the study of the long-term effect of motion types (forward, sideways, and backward walking) on the localisation accuracy. Most of the body-worn IMU solutions are designed for forward walking, such as SHS-based IMU localisation methods (describe in Section 2.1.2), and only work in forward walk direction. If the subject walks in a sideways or backward direction, the algorithm still assumes forward motion, which leads to positioning inaccuracies. Even though the foot-mounted IMU's ZUPT algorithm, the method used by our system, detects and registers backward and sideways walks, the positioning accuracy is not the same as the forward walk [196, 197]. Some research work has adapted the ZUPT algorithm to work for different walking speeds [198]. However, there has not yet been any study isolating the sideways and backward walk from the forward walk and measuring the long-term effect on positioning accuracy. Apart from these three walking motion types, first responders movement also consists of: running, climbing, jogging, crawling, standing, jumping, shuffling, crouching, crouch walking, and knee-dragging motion types [14, 15, 55, 199–202]. Results in the past research indicated that each motion type led to a different localisation accuracy [15, 201, 203]. To find a solution that works for all first responder movements, these motion types should be studied individually and in combination. To make the localisation solution more tractable, we focused our research on three walking motion types which could be expanded to other motion types in the future research.

5.2 Distance-based Error Correction Approach

Our proposed distance-based error correction approach used the localisation system developed in Chapter 4 by loosely coupling the foot-mounted IMU and camera-based system. However, in the distance-based error correction approach, instead of keeping the camera always on, its operation was modulated using the distance estimated from

the foot-mounted IMU with the localisation algorithm explained in Section 4.1. Figure 5.1 illustrates how we implemented our proposed method. The localisation system was initialised by the visual sensor which then went into sleep mode. From this point onwards, the localisation information of the tracked person was updated by the foot-mounted IMU until the distance threshold was reached. The distance estimated by the foot-mounted IMU system was compared against the preset distance threshold, and this comparison was a determining parameter to trigger the correction mechanism from the camera. The first distance threshold is relative to the starting point, and the subsequent thresholds are relative to the next correction point. The flowchart in Figure 5.2 shows the steps involved in the implementation of the approach. Once the distance threshold was reached, the camera system was set to the awake state to capture images of fiducial markers for 1.0 second to discretely correct the error accumulated by the foot-mounted IMU - the default time of 1.0 second was empirically determined through experimentation to be enough for camera pose estimation when markers are in camera's field of view. Once the correction was applied, the camera system went back into sleep mode, and the foot-mounted IMU started updating the position and orientation of the tracked person from the corrected localisation values. In case of visual localisation failure, re-localisation is attempted at the next stance phase with camera turning on for 1 second until the correction is achieved, and the next distance threshold is set from the corrected point. With this approach, we were able to reduce the use of the camera system and still correct the accumulated localisation error from the foot-mounted IMU. Through experimentation we found $25m$ to be an appropriate distance threshold. Therefore, we chose the $25m$, $50m$, $75m$, and $100m$ distances for the experiment carried out in Section 5.4.

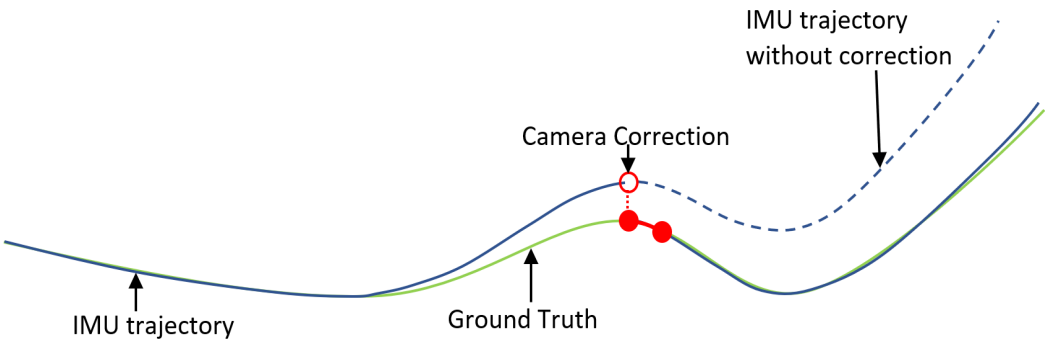


FIGURE 5.1: Correction of the Foot-mounted IMU trajectory with the camera-based localisation system

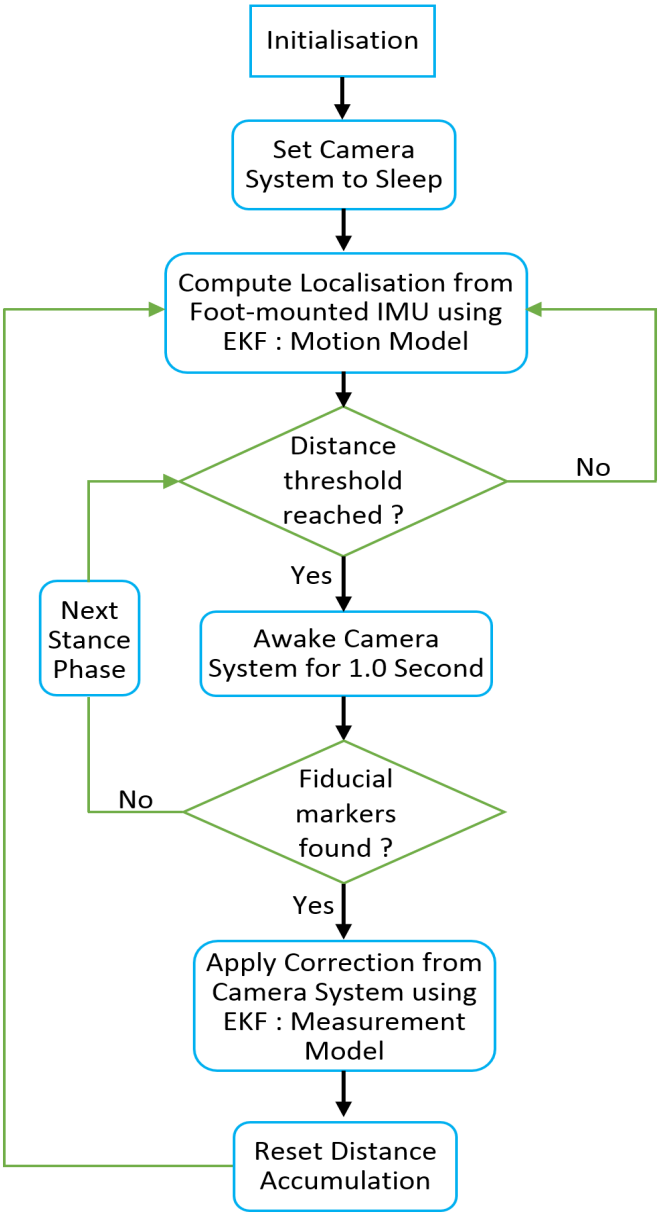


FIGURE 5.2: Flowchart of the distance-based error correction approach

5.3 Measuring Energy Consumption

Accurately measuring the energy consumed is critical to evaluate the localisation system for energy consumption and understanding the battery run time. Measuring energy consumption of a system involves three components: (1) the voltage drop, (2) the current flowing, and (3) the time. The voltage of the system is directly measured by the multimeter; however, the current is indirectly calculated by Ohm's law. The multimeter has a known shunt resistor, and measures the voltage drop across it. The current in

the circuit is equal to the voltage divided by the shunt resistance value. The voltage drop across the shunt resistor is known as the burden voltage, and it increases with the current, typically specified in mV/A. If the burden voltage is comparable to the voltage drop across the measured device, it leads to a higher error in current measurement which in turn means a wrong energy consumption value. Therefore, the shunt resistor should be changed to reduce the burden voltage and give higher resolution of the measured current value. Most multimeters have multiple shunt resistors, however, they have to be changed manually by plugging the connecting leads to a different input. In our case, when the camera goes from an awake to a sleep state, there is a significant change in current range from hundreds to tens of milliamps. This change requires dynamic adjustment of the shunt resistance which cannot be done by multimeters. There are only three devices to the best of our knowledge which minimise the burden on voltage and dynamically change the shunt resistance within the consumer budget. These devices are: (1) joulescope [204], (2) Otii DC power analyzer [205], and (3) current ranger [206]. We chose Joulescope as it also has a USB power input that allows direct interfacing with the camera and IMU system.

To measure the current and the voltage of the camera and the IMU system, we created a bench setup with the joulescope. The setup is shown in Figure 5.3. Since it is not possible to measure the energy consumption of the system while walking the $250m$ trajectory, a long bench test was performed on each component of the localisation system. The camera used for our system, Logitech C920, was set to the same settings as the localisation experiments. The video resolution was configured to 1920x1080, the white balance to auto, and the focus to manual mode. The bench test was 10 minutes for the Logitech C920 camera and the IMU wireless dongle. The XSens MTW Awinda IMU is powered through an internal battery, so the energy consumption was indirectly measured by monitoring the battery level while charging and discharging. The charging was for 3, 6, and 9 minutes, and the discharging test was 1 hour long. The 10 minute bench test consisted of 5 minutes of an awake state and 5 minutes of a sleep state. Each component was tested separately with the similar conditions as the experiment. During the camera test, the camera was looking at the fiducial markers and recording the video. For the IMU dongle test, the XSens IMU was connected to the dongle. In the case of XSens MTW Awinda IMU (shown in Figure 5.4), the IMU was attached to a mobile phone and mounted on a swinging motion device to keep the IMU oscillating similar to the walking

motion. This results in a similar energy consumption as during walking. Mobile phone was just used to attach the IMU, and it can be replaced with anything (such as wooden block) with matching physical dimensions.

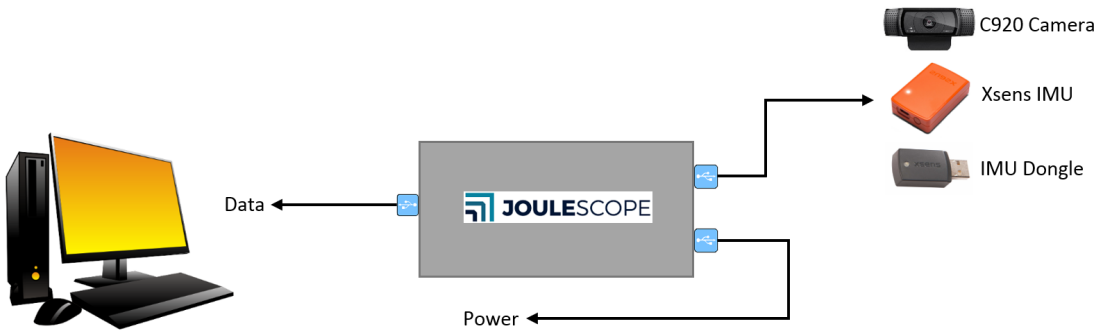


FIGURE 5.3: Joulescope bench test setup

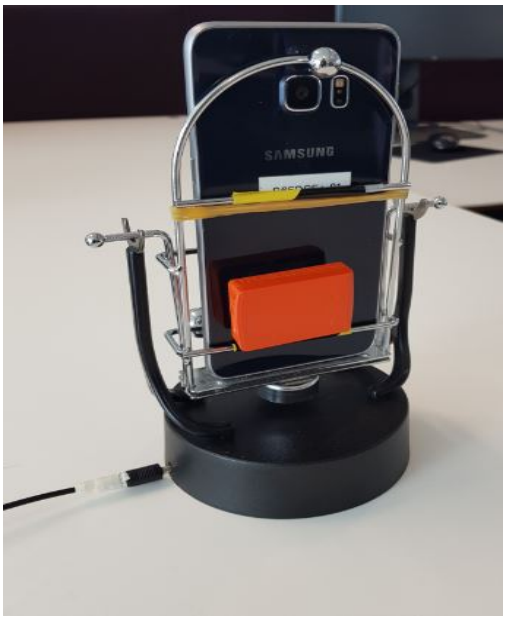


FIGURE 5.4: XSens MTW Awinda IMU attached to the swinging motion type device

The results of the Logitech C920 camera's bench test are in Table 5.1, and the dongle results are in Table 5.2. As our energy calculation involved multiplication of power and time, so we used 1 J as the energy unit in this thesis, which is equal 1 joule (J) of energy. The IMU charging was done at 3, 6, and 9 minute intervals. During charging, the battery level increase was 5% for 3 minutes, 9% for 6 minutes, and 14% for 9 minutes. Mean recharged power was 730.71 mW, and the charging current and voltage were 150.24 mA and 4.86 V, respectively. From the charging results, we calculated the 1% change in battery level which took 38.19 second on average while using 7.75 mWh or 27.9 J energy. The IMU's discharging was done on the phone swinging motion platform. There

were three repetitions of the experiment of one hour each. In the first repetition, the battery level decreased from 26% to 13%. For the second repetition, it decreased from 38% to 25%, and in the third repetition, decreased from 32% to 19%. For 1% decrease in battery level, the mean time taken was 4 minutes and 39.87 seconds. The current and voltage measured during recharging was then used to calculate energy consumed during 1% discharging, which is 7.75 mWh.

TABLE 5.1: Camera bench test results for sleep and awake state

Camera State	**Current**(mA)	**Voltage**(V)	**Power**(mW)	**Time** (min)
Sleep	47.773	5.041	240.8	5
Awake	215.44	4.975	1071.7	5

TABLE 5.2: IMU dongle bench test results for sleep and awake state

IMU State	**Current**(mA)	**Voltage**(V)	**Power**(mW)	**Time** (min)
Sleep	25.457	5.038	128.25	5
Awake	32.962	5.031	165.83	5

The camera was also tested for a 1.0 second sleep-to-awake and awake-to-sleep state transition. The current, voltage, and power plots from the test are shown in Figure 5.5. Markers identified during this test can be seen in Figure 5.6. The sleep-to-awake transition time was 0.2 seconds with a power consumption of 730 mW, whereas the awake-to-sleep transition time was 0.53 seconds with a power consumption of 75 mW. The transition time and power consumption were later applied to the energy consumption calculation during experiments along with the awake state power consumption values. The current and power signals in Figure 5.5 also show the number of frames transmitted from the camera to computer through Joulescope, where each frame has a transmission signal outline as indicated on Figure 5.5.

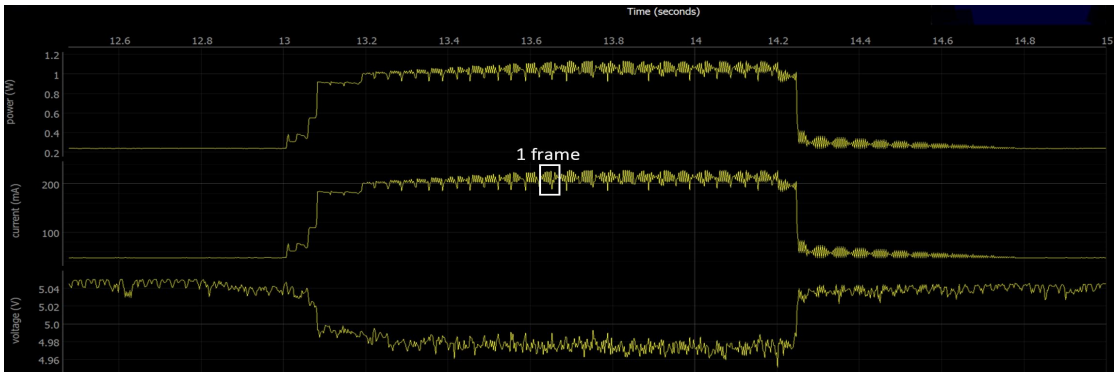


FIGURE 5.5: Joulescope's power, current, and voltage plots for the 1.0 second video

FIGURE 5.6: Marker identification during the 1.0 second video

5.4 Experiment

The experiment consisted of three conditions: forward walk, sideways walk, and backward walk. For each motion type, five loops were done of the planned trajectory shown in Figure 4.9, resulting in a $250m$ long walking trajectory. Each motion type was repeated three times. During each repetition, data from the foot-mounted IMU and the camera-based system was collected on the Razer Blade laptop and processed afterwards. The experiment was carried out by a healthy male individual walking at a normal pace (nearly $1.4m$ per second), who took 5-10 minute break between each repetition. The forward walk took less time to complete than the sideways and backward walks, and it was also least physically demanding. The sideways walk was the most physically straining, and required a 10 minute rest before the next repetition. The backward walk was not physically challenging, but difficult to perform as the subject was facing an opposing view to the walking direction which made it harder.

5.4.1 Results

a) Forward Walk

The results of the three forward walking trajectories of the foot-mounted IMU are shown in Figure 5.7. The eCDF values are also computed for the three uncorrected trajectories. The localisation errors for Log1, Log2, and Log3 are: $1.58m$, $1.34m$, and $1.65m$ for 95% of the $250m$ walk, respectively. The resulting trajectories of the forward walk for Log 1, Log2, and Log3 are quite similar, so in order to save space, we chose Log 1 and its corrected trajectories to represent forward walk in this thesis. For Log 1, the distance-based camera correction was applied at $25m$, $50m$, $75m$, and $100m$ intervals. During the correction, the absolute position and orientation values were computed from the video frames captured in 1.0 second. For the $25m$ distance threshold, the first correction from the camera system was applied at $25m$ of the distance estimated by the foot-mounted IMU; the second was at $50m$, then $75m$, $100m$, $125m$ and so on. Similarly, the distance-based camera correction method was used for $50m$, $75m$, $100m$ distance thresholds. There were 9 corrections for $25m$ distance threshold, 4 for $50m$, 3 for $75m$, and 2 for $100m$ distance threshold. The estimated energy spent on each correction was 1.257 J. The resulting trajectories of these corrections are shown in Figure 5.8, and the eCDF percentile versus the localisation error plots for these trajectories are in Figure 5.9. The time taken to complete the forward walk for Log 1 was 4 minutes and 35.0 seconds. Table 5.3 shows the localisation error at 95th percentile and energy consumption of the uncorrected and corrected trajectories.

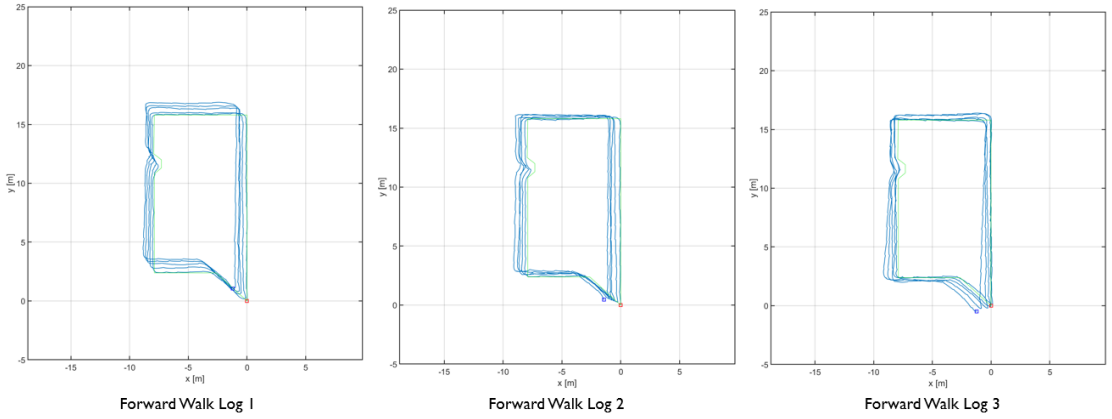


FIGURE 5.7: Forward walk trajectories of the foot-mounted IMU

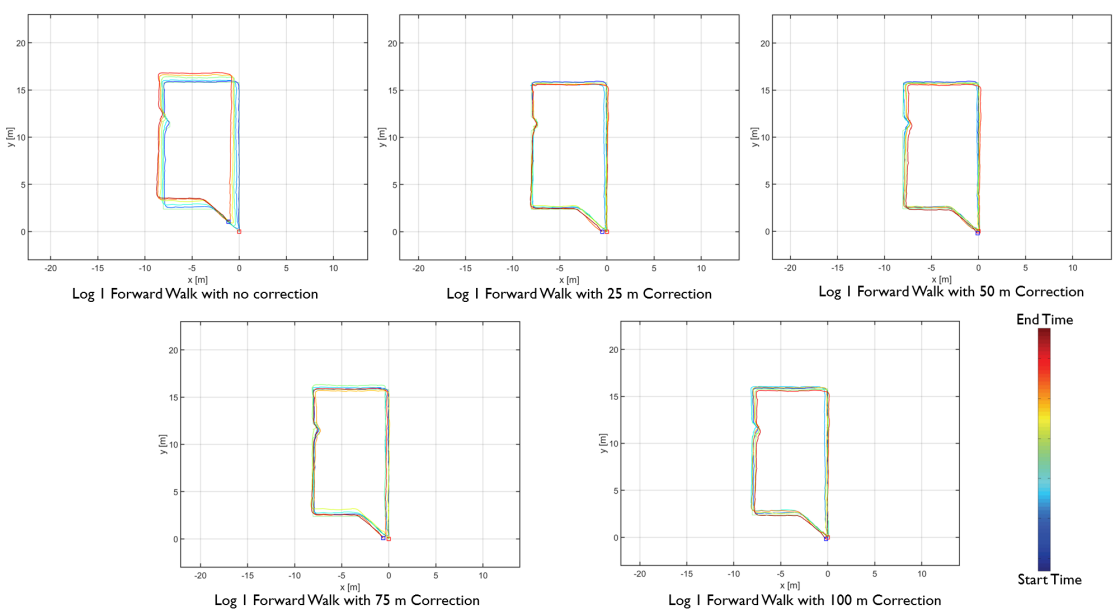


FIGURE 5.8: Forward walk trajectories with the distance-based camera correction at $25m$, $50m$, $75m$, and $100m$ intervals

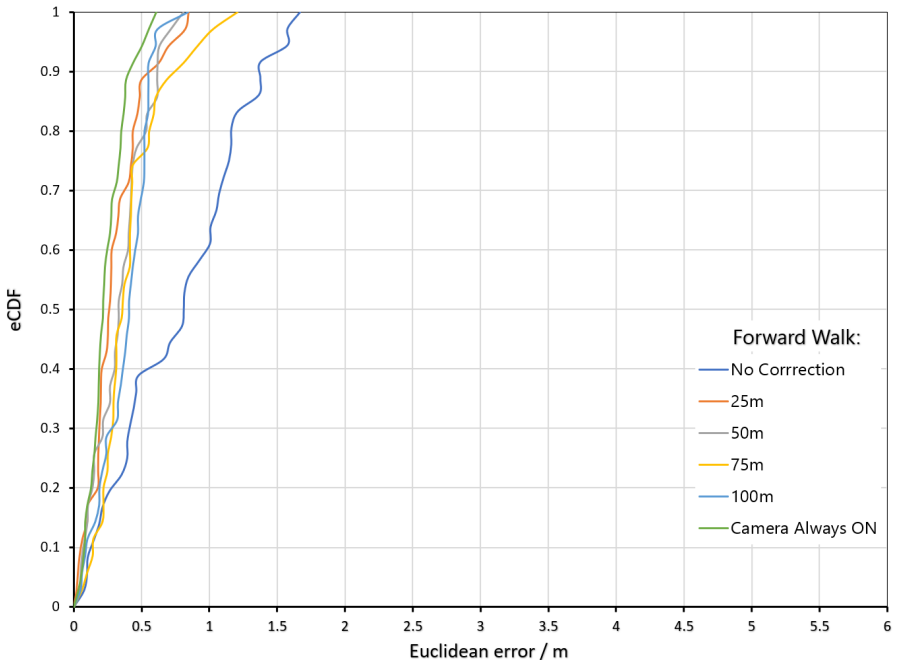


FIGURE 5.9: Localisation error (m) versus eCDF percentile for forward walk

TABLE 5.3: Forward walk - energy consumed (J) by each device and the localisation error (m) at 95th percentile for uncorrected and distance-based corrected trajectories

Correction Type	**Camera** (J)	**IMU** (J)	**Dongle** (J)	**error** (m)
Camera Always ON	294.7	27.4	45.6	0.52
25m Correction	32.4	27.4	45.6	0.61
50m Correction	26.1	27.4	45.6	0.63
75m Correction	24.9	27.4	45.6	0.72
100m Correction	23.6	27.4	45.6	0.75

b) Backward Walk

Figure 5.10 shows the results of three backward walking foot-mounted IMU trajectories. Similar to forward walk, we computed the eCDF value of the localisation error. Log1, Log2, and Log3 of the backward walk have 3.13m, 2.45m, and 1.82m localisation errors for 95% of the 250m walk, respectively. Same as forward walk, we chose Log 1 of the backward walk and its corrected trajectories to represent backward walk in this thesis. Figure 5.11 shows the sensor fusion plots of Log1 with the distance-based error correction method applied with 25m, 50m, 75m, and 100m distance thresholds. The eCDF percentile versus the localisation error plot of the uncorrected and corrected trajectories is in Figure 5.12. The Log 1 of the backward walk took 8 minutes and 50 seconds of walking to complete the 250m distance. Table 5.4 shows the localisation error at the 95th percentile and energy consumption of the uncorrected and corrected trajectories.

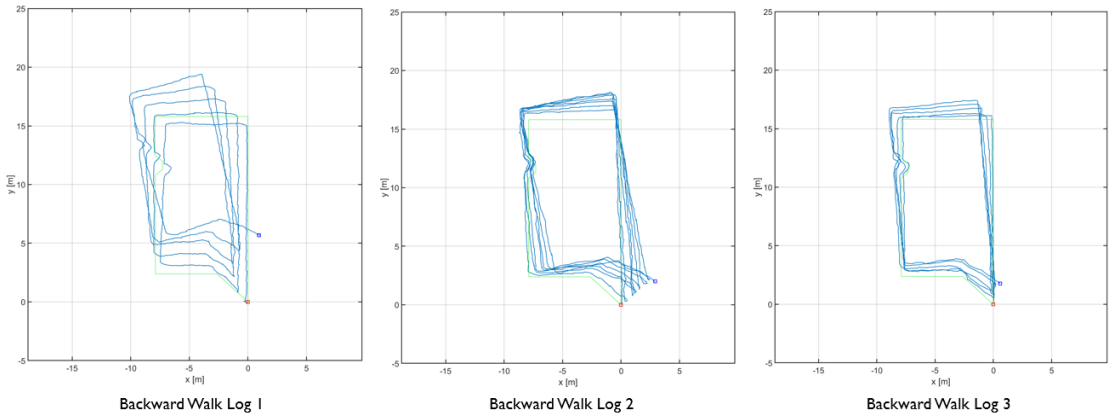


FIGURE 5.10: Backward walk trajectories of the foot-mounted IMU

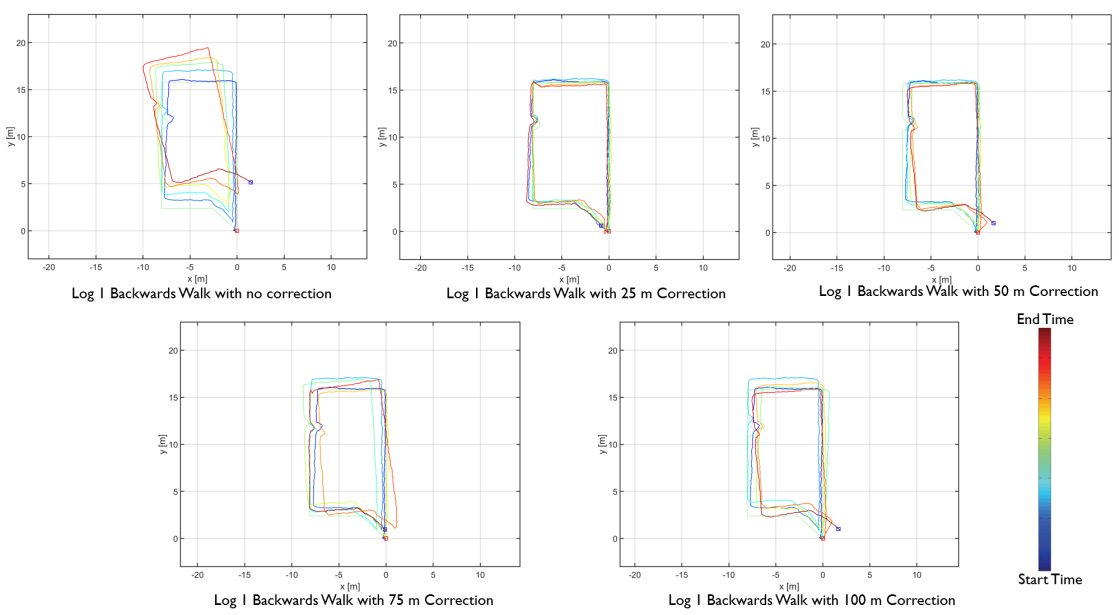


FIGURE 5.11: Backward walk trajectories with the distance-based camera correction at $25m$, $50m$, $75m$, and $100m$ intervals

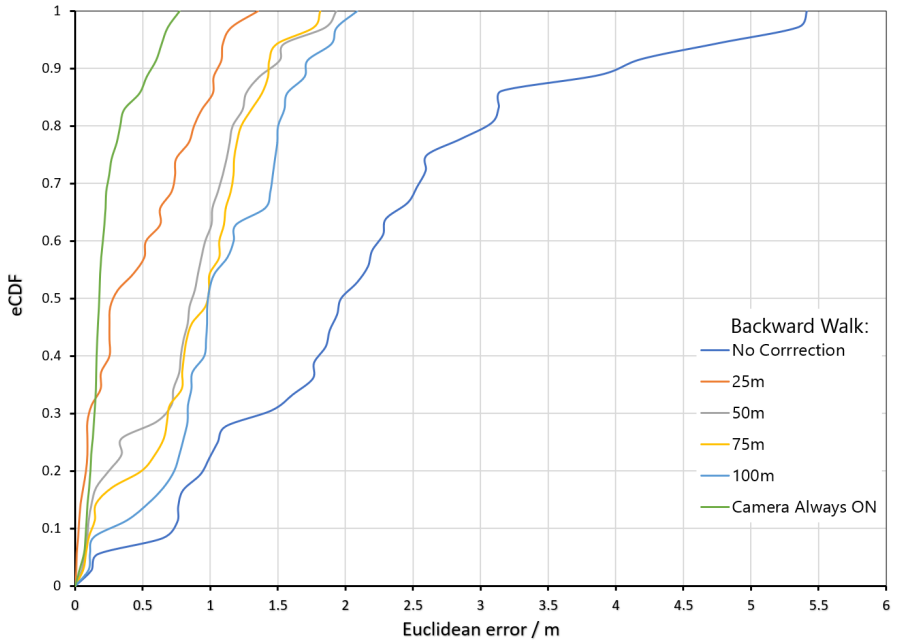


FIGURE 5.12: Localisation error (m) versus eCDF percentile for backward walk

TABLE 5.4: Backward walk - energy consumed (J) by each device and the localisation error (m) at 95th percentile for uncorrected and distance-based corrected trajectories

Correction Type	**Camera** (J)	**IMU** (J)	**Dongle** (J)	**error** (m)
Camera Always ON	568.0	52.8	87.9	0.65
25m Correction	32.4	52.8	87.9	1.10
50m Correction	26.1	52.8	87.9	1.55
75m Correction	24.9	52.8	87.9	1.72
100m Correction	23.6	52.8	87.9	1.90

c) Sideways Walk

The results of the three sideways walking foot-mounted IMU trajectories are shown in Figure 5.7. The localisation error for Log1, Log2, and Log 3 at 95% of the walk are 9.63m, 6.49m, and 7.54m, respectively. Similar to forward and backward walks, Log1 and its corrected trajectories were chosen to represent sideways walk in this thesis. Sensor fusion plots for Log1 with the distance-based error correction with 25m, 50m, 75m, and 100m distance thresholds are shown in Figure 5.14. The eCDF percentile versus the localisation error plot of the uncorrected and corrected trajectories is in Figure 5.15. The Log 1 of the sideways walk took 8 minutes and 10 seconds of walking to complete the 250m distance. Table 5.5 shows the localisation error at the 95th percentile and energy consumption of the uncorrected and corrected trajectories.

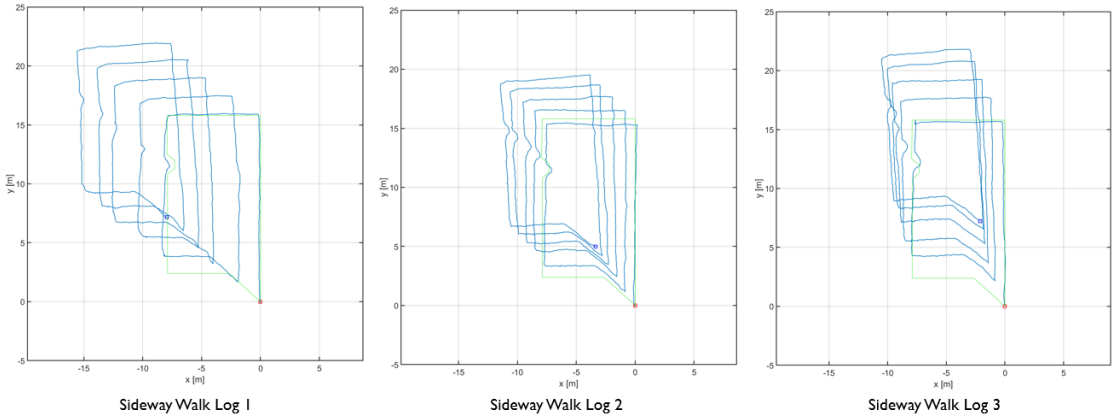


FIGURE 5.13: Sideways walk trajectories of the foot-mounted IMU

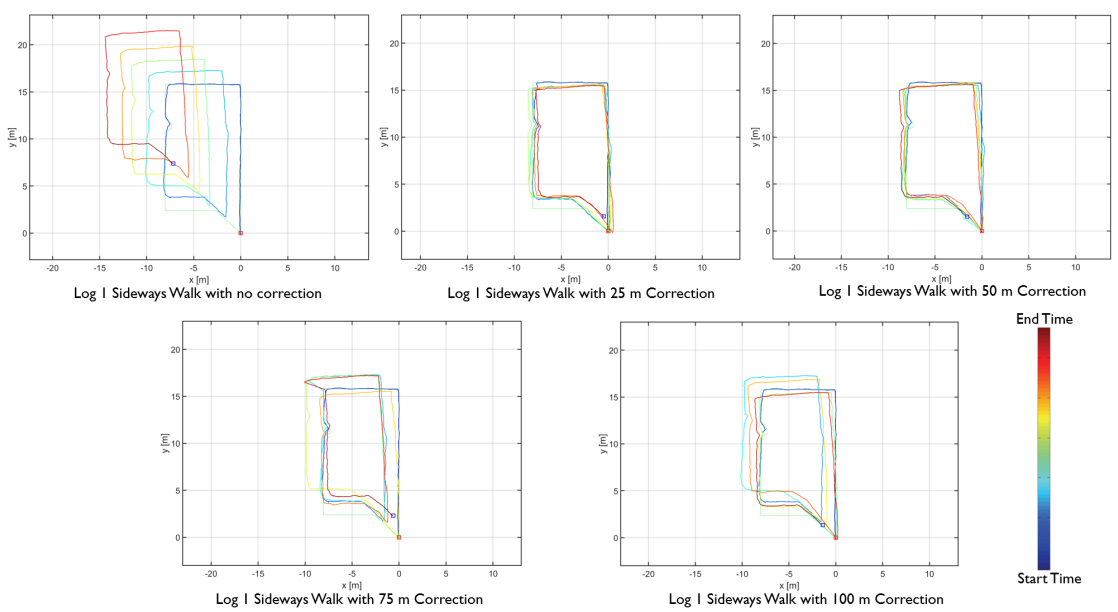


FIGURE 5.14: Sideways walk trajectories with the distance-based camera correction at $25m$, $50m$, $75m$, and $100m$ intervals

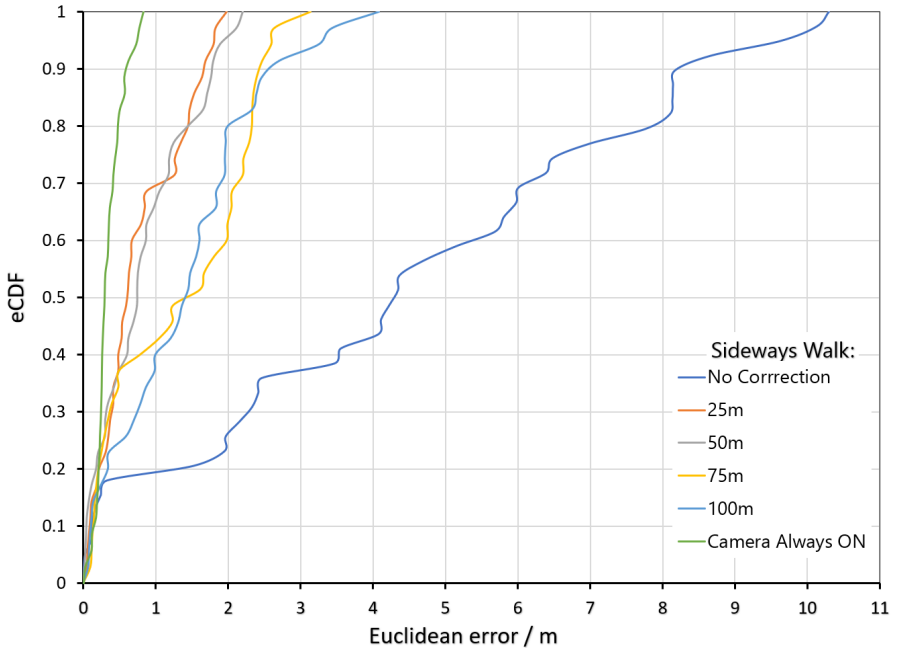


FIGURE 5.15: Localisation error (m) versus eCDF percentile for sideways walk

TABLE 5.5: Sideways walk - energy consumed (J) by each device and the localisation error (m) at 95th percentile for uncorrected and distance-based corrected trajectories

Correction Type	**Camera (J)**	**IMU (J)**	**Dongle (J)**	**error (m)**
Camera Always ON	525.1	48.8	81.3	0.78
25m Correction	32.4	48.8	81.3	1.80
50m Correction	26.1	48.8	81.3	1.95
75m Correction	24.9	48.8	81.3	2.60
100m Correction	23.6	48.8	81.3	3.31

5.4.2 Discussion

The experimental results of the distance-based camera correction approach, Tables 5.3, 5.4, and 5.5, show that by reducing the usage of the camera system, the energy consumption of the localisation system decreases by more than 70%. The results are also verified by the energy/accuracy trade-off plots shown in Figures 5.16, 5.17, and 5.18. The highest reduction in energy consumption occurred when the localisation system switched from continuous (Camera always ON) to a discrete distance-based camera correction approach. As the number of corrections between the $25m$, $50m$, $75m$, and $100m$ distance thresholds were similar, we only observed smaller decreases in energy consumption. For localisation accuracy, unlike the energy consumption, a gradual decrease was observed as can be seen in Figures 5.16, 5.17, and 5.18. Even after the decrease, the localisation accuracy still remained within the required sub-metre range for the forward walk motion type. For the backward and sideways walk motion types, the localisation accuracy fell below the sub-metre range. However, in real-life scenarios, first responder motion consists of mixed motion types, and it is unlikely that they would walk for more than 10 to 15 metres in sideways or backwards motion. The mixed motion type scenarios with short distance sideways and backward walks are covered in Chapter 6. With the results obtained, we can see the advantages of using the discrete distance-based camera correction approach. The method would enable long-term first responder localisation during training exercises without recharging batteries. The results also suggest that near $25m$ distance corrections would ensure sub-metre localisation accuracy if the walk mostly consists of the forward motion type.

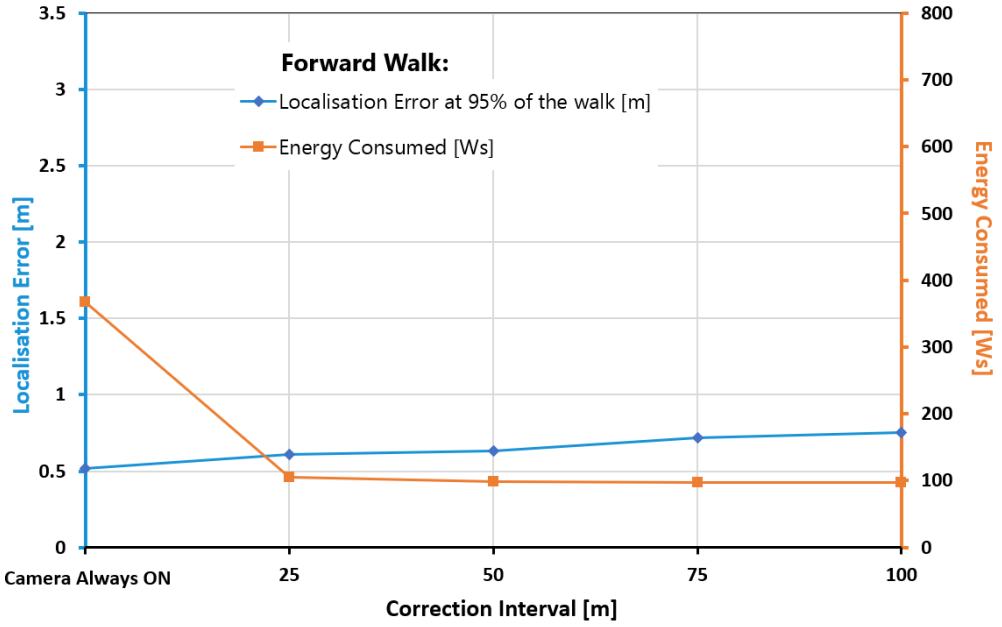


FIGURE 5.16: Energy/accuracy trade-off plot for forward walk motion type

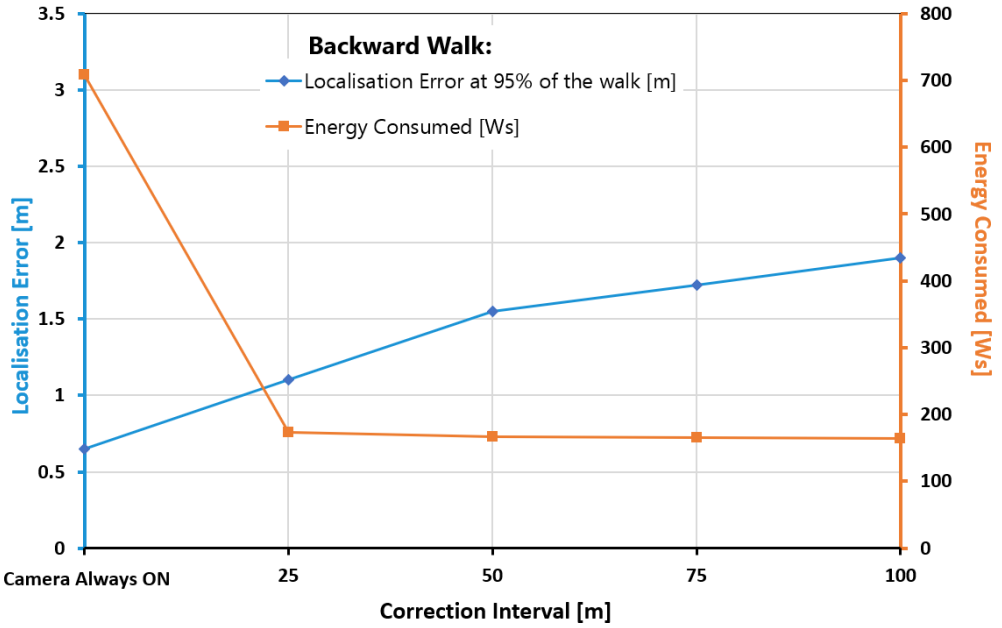


FIGURE 5.17: Energy/accuracy trade-off plot for backward walk motion type

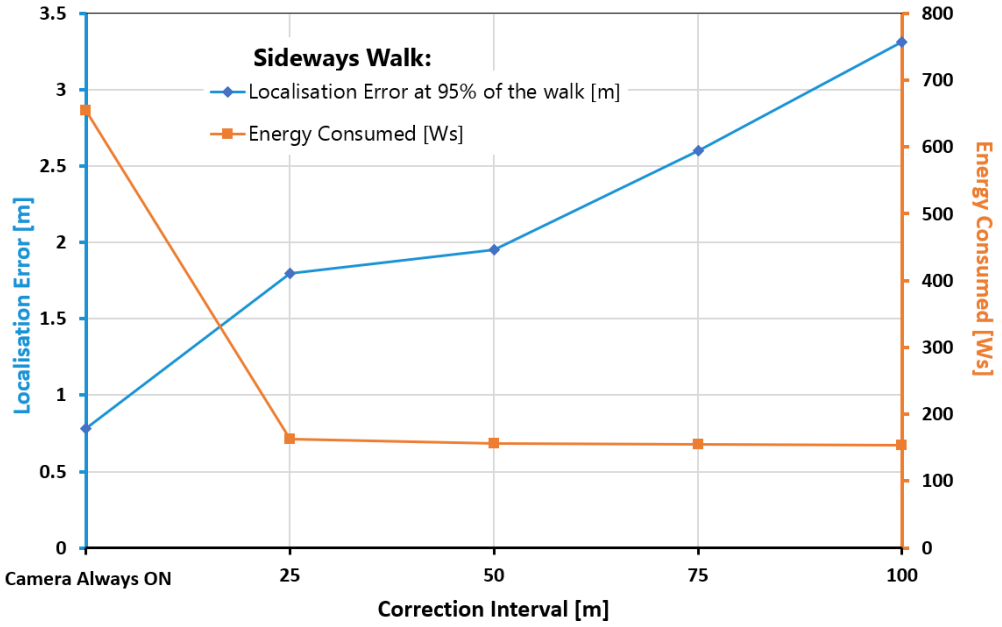


FIGURE 5.18: Energy/accuracy trade-off plot for sideways walk motion type

To understand the long-term effect of motion types on the localisation accuracy, we analysed the uncorrected forward, backward, and sideways walking motion type trajectories, which are shown in Figures 5.7, 5.10, and 5.13. As expected, the forward walk motion type was the most accurate of the three with a mean localisation error of 1.52 m for 95% of the walk. The backward walk was the next with a mean localisation error of $2.48m$ for 95% of the walk. On observing the three backward walk trajectories (Figure 5.10), a constant drift in heading values can be seen in all three trajectories, which leads to a decrease in the overall localisation accuracy. Least in accuracy is the sideways walk motion type with a mean localisation error of $7.89m$ for 95% of the walk. All three trajectories of sideways walk (Figure 5.13) show a large scaling error, even though the heading values are accurate. However, the scaling error has a notable cumulative effect on the subsequent position and heading values causing a gradual increase in the localisation error. The large errors in sideways and backward walks can be explained when we plot the accelerometer's acceleration and gyroscope's angular velocity data. Each motion type exhibits a unique acceleration and angular velocity signal, shown in Figures 5.19, 5.20, and 5.21. The distinct nature of these signals can be used to segment and classify motion types. A more in-depth analysis of the signals, such as computing the velocity and angular signals, could also reveal more details on the cause of localisation error. This could be pursued in future research.

Another approach which could be explored in future research is dynamically adjusting the distance threshold. There are four key factors effecting the localisation accuracy of the foot-mounted IMU: motion types (walking, running, crawling etc.), gait speed, walking surface, and mounting position of the IMU (toe, heel, or ankle) [15]. A more dynamic approach considering all or some of these factors and varying the correction mechanism would further improve the energy efficiency of the localisation system. However, the distance travelled with a certain motion type, gait speed, and walking surface would still need to be accounted for in the dynamic approach to vary the integrated distance threshold with the change in each factor.

FIGURE 5.19: Forward walk's accelerometer and gyroscope signals

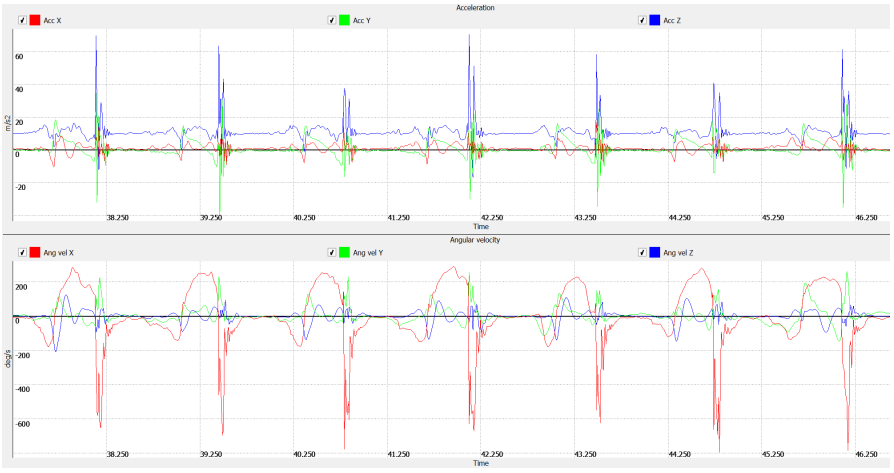


FIGURE 5.20: Backward walk's accelerometer and gyroscope signals

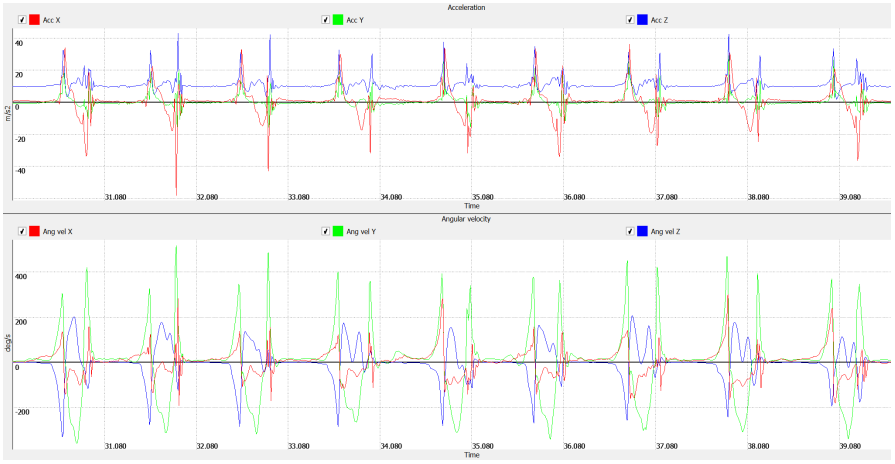


FIGURE 5.21: Sideways walk's accelerometer and gyroscope signals

5.5 Conclusion

In this chapter, we devised and evaluated a distance-based error correction approach to increase the energy efficiency of the localisation system. With our proposed approach, we were able to decrease the energy consumption of the localisation system by more than 70% for three different walking motion types (forward, backward, and sideways walks). The highest reduction in energy consumption occurred when the localisation system switched from continuous to discrete distance-based camera correction. As the number of corrections between the $25m$, $50m$, $75m$, and $100m$ distance thresholds were similar, we only observed smaller decreases in energy consumption. The localisation accuracy between the $25m$, $50m$, $75m$, and $100m$ distance thresholds differed largely, with $25m$ being the most accurate. We also studied the long-term effect of forward, backward, and sideways walking motion types on localisation accuracy. The mean localisation errors for forward, backward, and sideways walks were $1.52m$, $2.48m$, and $7.89m$, respectively, for 95% of the walk. We also discovered that each motion type has unique acceleration and angular velocity signals which can be used to segment the motion type. With the segmented trajectory, a different correction mechanism could be used for each motion type to reduce the localisation error. This would be particularly useful in real-life scenarios where walking consists of mix motion types. This approach is further investigated in Chapter 6.

Chapter 6

Motion Type Adaptive Positioning Algorithm

In the last chapter, a long-term study on walking motion types revealed the variation in localisation accuracy for different motion types when using foot-mounted IMU with the ZUPT algorithm. We learnt that applying discrete distance-based error correction from the camera system decreases the energy consumption of the localisation system by more than 70%, while keeping the accuracy to a near sub-metre level. The experiment conducted in Chapter 5 consisted of either forward, backward, or sideways walking motion types. However, in a real-life, walking can vary between different motion types (such as forward to backward, then back to forward or sideways) especially in training scenarios. This chapter presents the research on these variations and derives a correction mechanism based on the error accumulation rate of each motion type. As we observed in Section 5.4.1, the forward walk was the most accurate of the three walking motion types, indicating the least amount of error accumulation. In order to conserve energy while maintaining a higher accuracy, the localisation system should consider the variations and use them to change the correction mechanism. This chapter describes the implementation and evaluation of a novel motion type adaptive correction approach that determines the next correction from the camera system using the current walking motion type and required localisation accuracy.

The chapter starts with the review of past research on motion types and the ZUPT algorithm (Section 6.1). Then the chapter describes the proposed motion type adaptive

correction approach (Section 6.2). Subsequently, the classification approaches to distinguish between motion types based on the inertial sensor data are discussed (Section 6.3), followed by the section on the conducted experiment evaluating the proposed approach (Section 6.4), and lastly, the conclusion (Section 6.5).

6.1 Motion Types and the ZUPT Algorithm

A robust ZUPT algorithm that works accurately on a range of motion types and different speeds of motion is still an open research problem [15, 201, 203]. As discussed in Section 2.1.2, unlike the SHS-based PDR algorithm, ZUPT works on a variety of motion types, but its accuracy deteriorates when the motion type is not a forward walk. Even with these shortcomings, ZUPT is the most accurate algorithm to date for estimating human position and orientation from an IMU [163]. To mitigate these errors due to motion types, recent research has tuned ZUPT to work on distinctly dissimilar motion types [164, 184, 207]. However, the subtle variation in motion types, such as different types of walking or running, has not been studied in the past [208–210]. The focus of this chapter is to study the localisation error accumulation in walking motion types and use the results to provide adaptive correction to maintain the position accuracy within the required sub-metre range.

6.2 Motion Type Adaptive Correction Approach

We devised an error correction mechanism whereby each motion type had an associated error accumulation parameter which defined the rate of error accumulation in its position value. During walking, when the motion type changed from one to the other, the error parameter changed as well. The flow chart in Figure 6.1 illustrates the steps involved in the implementation. Before the parameter values were applied, the walking motion type had to be recognised and classified into forward, backward, or sideways motion type. The classification process is explained in Section 6.3. After the motion type was recognised, the distance travelled in the current motion type and the associated error parameter were used to compute an estimate of the updated current accumulated error in position. This process was repeated on each detected step of the tracked person. The accumulated error was compared with the correction threshold. If the accumulated

error in the position value was equal to or exceeded the threshold value, the camera system was set to awake state and position and orientation correction was applied. The relationship between the error parameter and the localisation error correction threshold is given by:

$$T_{cp} \leq e_{fp} * d_f + e_{bp} * d_b + e_{sp} * d_s \tag{6.1}$$

where T_{cp}: Position correction threshold,

e_{fp}: Error accumulation parameter for forward walking,

d_f: Distance covered during the forward walking motion type,

e_{bp}: Error accumulation parameter for backward walking,

d_b: Distance covered during the backward walking motion type,

e_{sp}: Error accumulation parameter for sideways walking,

d_s: Distance covered during the sideways walking motion type

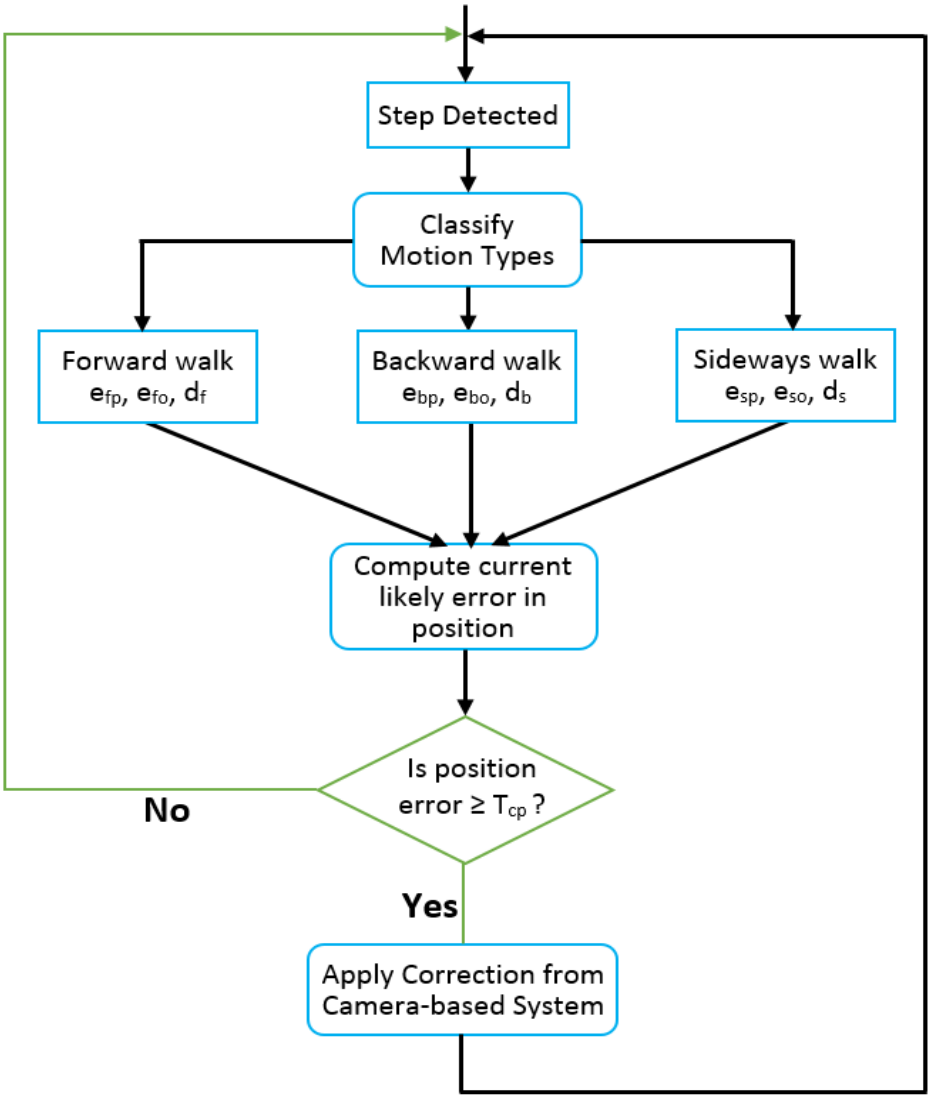


FIGURE 6.1: Flowchart of the motion type adaptive correction approach.

6.2.1 Data Collection and Computing Error Parameters

To assess the effectiveness of our idea, we recorded 15 trajectories of each motion type to calculate the position error parameter values, leading to a total of 45 walked trajectories for the three motion types. The data was collected using the foot-mounted IMU shown in Figure 4.2. The error parameters of each motion type were calculated by using the estimated position value by the ZUPT algorithm and the ground truth values at $10m$ intervals for all the 15 trajectories of the motion type. In total there were 375 data sample points for each motion type from the 15 recorded trajectories, and the error parameters were given by the mean value of these points. The box plots in Figure 6.2 show the distribution of the sampled position values.

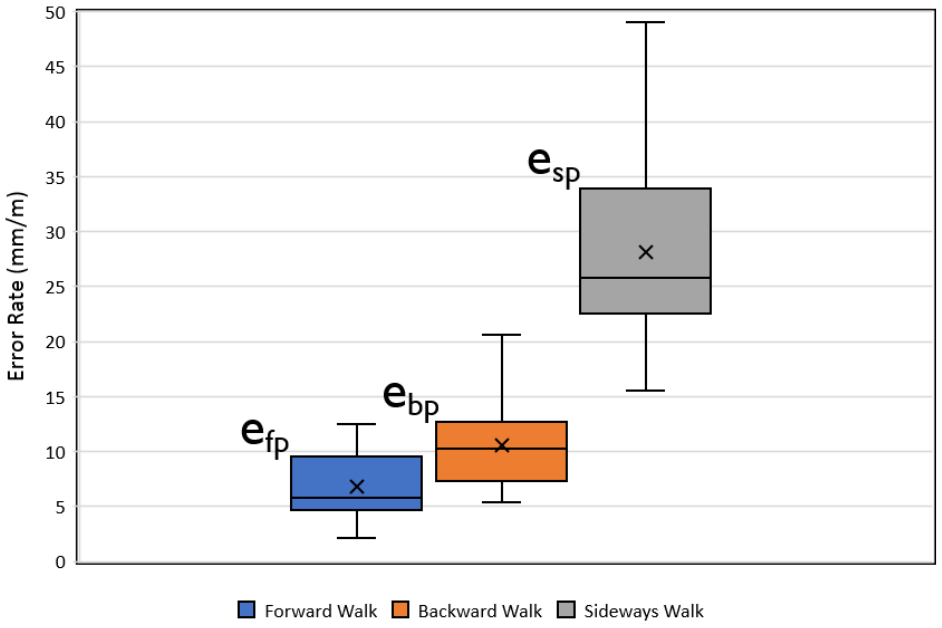


FIGURE 6.2: Box plots showing the distribution of position error parameters.

6.3 Classifying Motion Types

To apply the adaptive correction approach, the system should be able to distinguish between the motion types. There have been several classification methods proposed in past research to segment the motion types, primarily driven by wearable devices (smartwatches and fitness bands). Identifying motion types from the inertial sensor data involves three steps: (1) signal processing, (2) feature extraction, and (3) classification [133]. With signal processing, noise in the inertial sensor data is removed and signals are aggregated into analysable blocks using a process known as windowing. Each motion type has periodic acceleration and angular velocity signals as shown in Figures 5.19, 5.20, and 5.21, and windowing captures this periodicity for analysis. The second step for motion type recognition is the feature extraction. Features are the derived data values from the raw acceleration and angular velocity signals. The most commonly used features in the previous research are: **statistical features** (mean [211], variance [212], standard deviation [213], root mean square [214]), **time-domain features** (time between peaks [215], zero-crossing rate [216]), and **frequency-domain features** (Fourier transforms [214, 217], power spectral analysis [218], Wigner distribution [219]). The last step for motion type recognition is the classification from the extracted features. Classification methods used in the literature are: threshold analysis [220, 221], Bayesian

classifiers [222, 223], artificial neural networks [224, 225], decision trees [226, 227], hidden Markov model [221], support vector machine [213, 228], and regression methods [229, 230].

The classification accuracy to distinguish between motion types depends on the motion types present in the data and the three identification steps outlined above. If there is a clear distinction between the motion types, such as running, walking, lying down, crawling, and standing, the classification is going to be more accurate. However, if the distinction is subtle, such as different types of walking which is the case for our work, it becomes harder to distinguish. We have not yet found any solution focused on the variation in walking motion types (forward, backward, and sideways walks) [208–210, 231, 232]. Since the focus of this thesis is position accuracy and energy efficiency, we assumed a perfect classification for these motion types, and designed our experiment accordingly.

6.4 Experiment

To evaluate the motion type adaptive algorithm, we grouped the mixed motion types into groups of two and three motion types. In real-life scenarios, the backwards and sideways walking motion types last for shorter distances, so we kept the forward walk as a major part of the trajectory. Our planned walking trajectory for the group of two motion types consisted of $15m$ of the forward walk and $10m$ of the backward or sideways walk, and this pattern alternated for the entire $250m$ walk (five loops of the $50m$ trajectory shown in Figure 4.9). For the group of three motion types, our planned trajectory consisted of $15m$ forward, $10m$ backward, $15m$ forward, and $10m$ sideways walking motion types. Three repetitions were recorded for each combination of mixed motion types. The sensor setup for the experiment is shown in Figure 4.11. The data from the IMU and camera system was recorded on the Razer Blade laptop and processed afterwards. The experiment was carried out by a healthy male individual with a normal gait and walking at a normal pace of approximately $1.4m$ per second. Between each repetition, the participant took a 5 minute break.

6.4.1 Results

a) Forward and Backward Walk

The uncorrected and adaptively corrected walking trajectories of the forward and backward walking motion types are shown in Figure 6.3 and Figure 6.4, respectively. After applying the adaptive motion type correction, localisation errors were reduced to $0.88m$, $0.95m$, and $0.85m$ for 95% of the walk compared to $3.47m$, $3.50m$, and $3.41m$ for the uncorrected trajectories. To compare the performance of the adaptive correction with the distance-based correction method, we performed corrections with $20m$, $50m$, and $70m$ distance thresholds. The resulting trajectories from these corrections are shown in Figure 6.5, Figure 6.6, and Figure 6.7, respectively. To have a global view of the localisation error in three trajectories, eCDF values were combined into one plot for the uncorrected, adaptively corrected, and distance-based corrected trajectories. The resulting eCDF plots are shown in Figure 6.8.

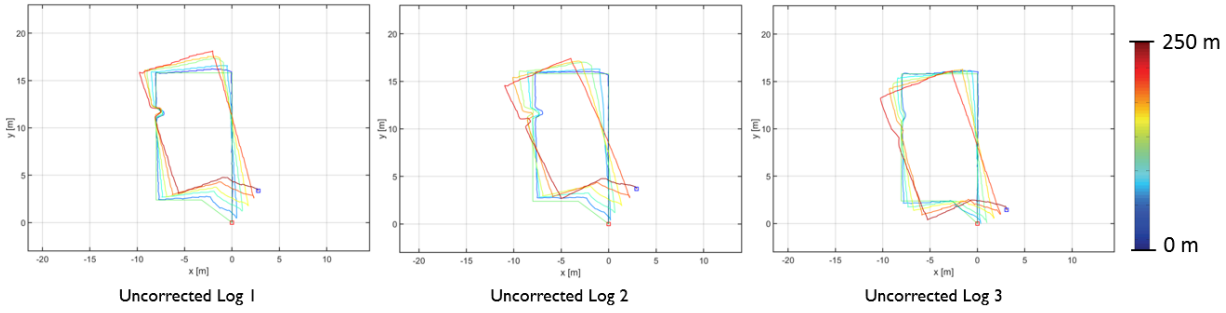


FIGURE 6.3: Uncorrected Forward and Backward walk trajectories.

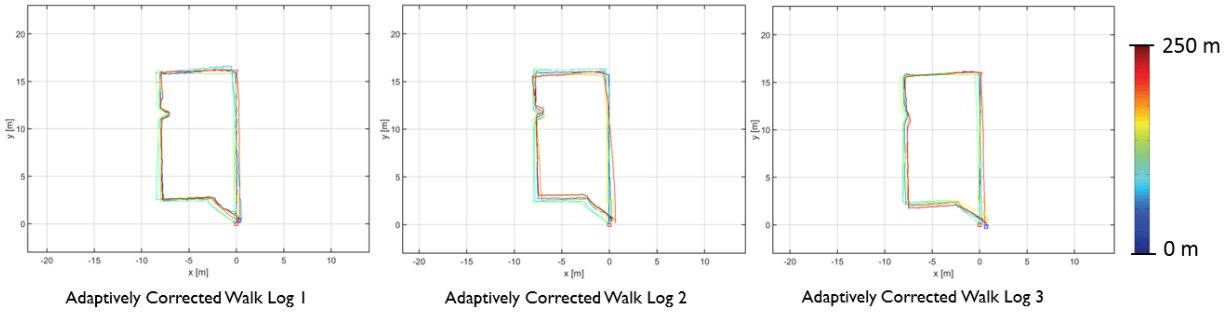


FIGURE 6.4: Adaptively Corrected Forward and Backward walk trajectories.

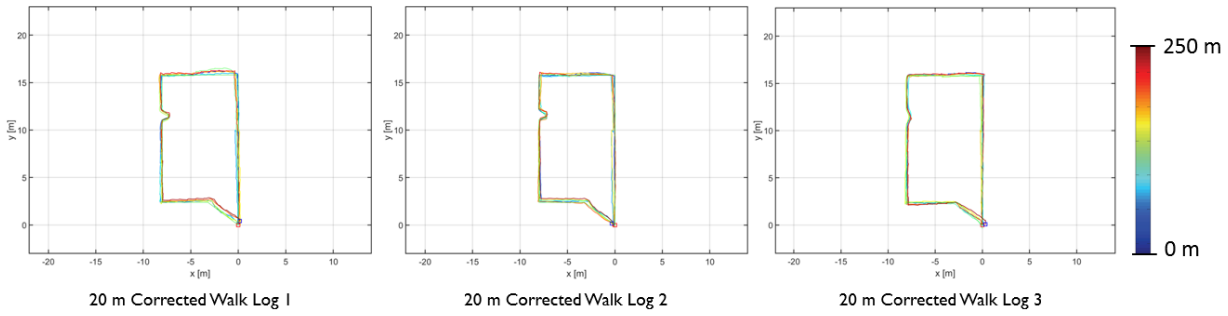


FIGURE 6.5: 20m Corrected Forward and Backward walk trajectories.

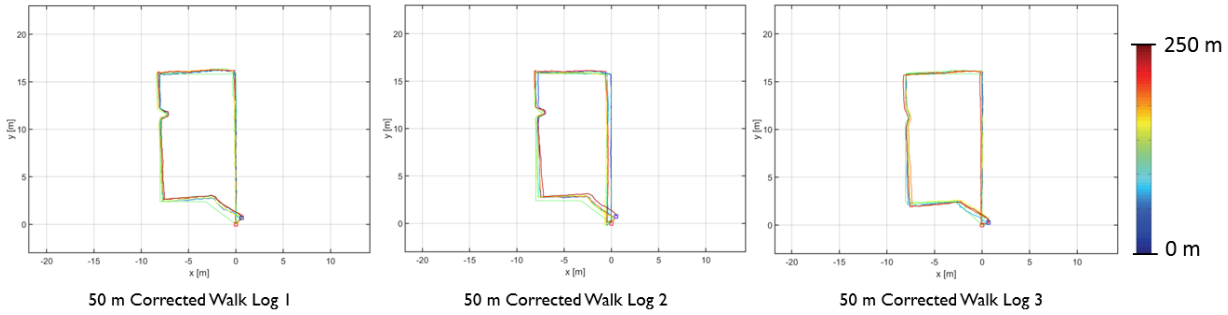


FIGURE 6.6: 50m Corrected Forward and Backward walk trajectories.

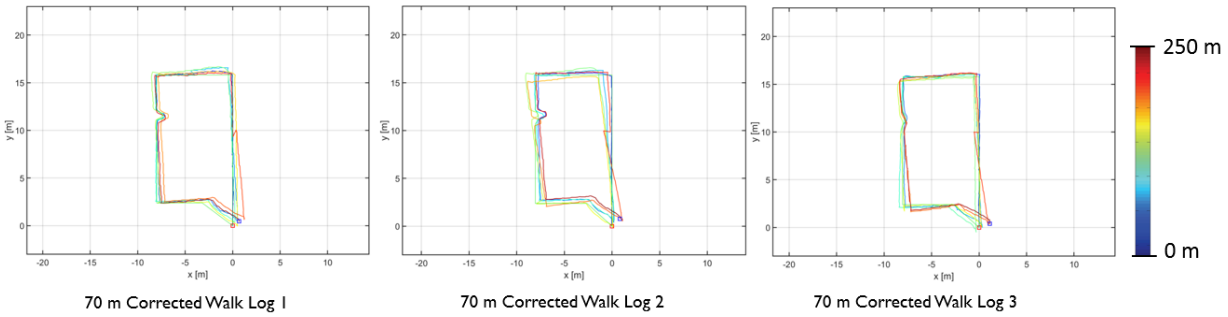


FIGURE 6.7: 70m Corrected Forward and Backward walk trajectories.

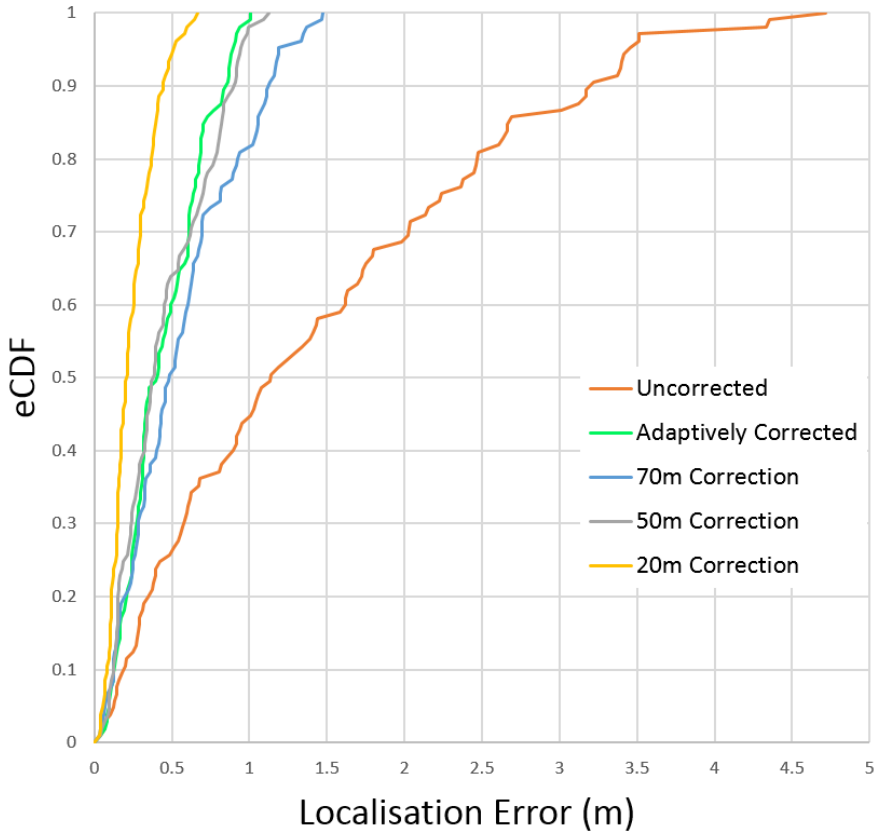


FIGURE 6.8: Combined eCDF values versus localisation error for the three trajectories.

b) Forward and Sideways Walk

For forward and sideways walk, the three uncorrected and adaptively corrected walking trajectories are shown in Figure 6.9 and Figure 6.10, respectively. The localisation errors for uncorrected trajectories were $3.37m$, $2.55m$, and $2.39m$ for 95% of the walk, and for adaptively corrected trajectories were $0.84m$, $0.80m$, and $0.76m$. For comparison with the distance-based correction method, the correction was applied at $20m$, $50m$ and $70m$ distance thresholds. The results are shown in Figure 6.11, Figure 6.12, and Figure 6.13, respectively. Combined eCDF values versus the localisation error for the three trajectories are plotted in Figure 6.14.

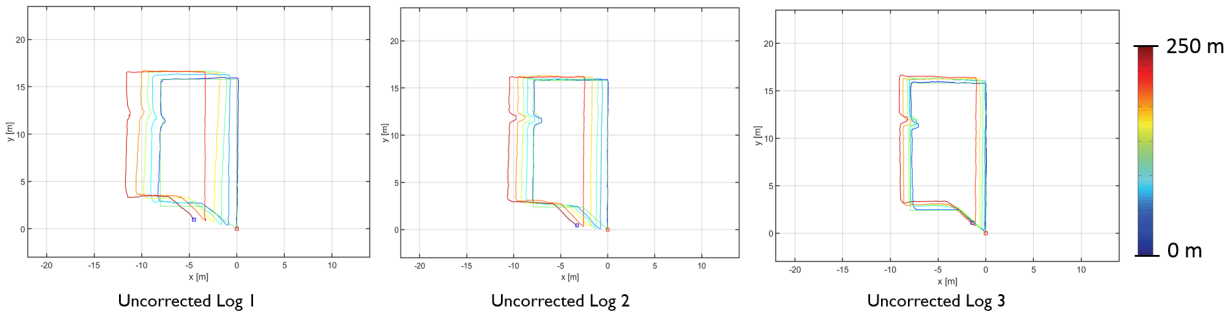


FIGURE 6.9: Uncorrected Forward and Sideways walk trajectories.

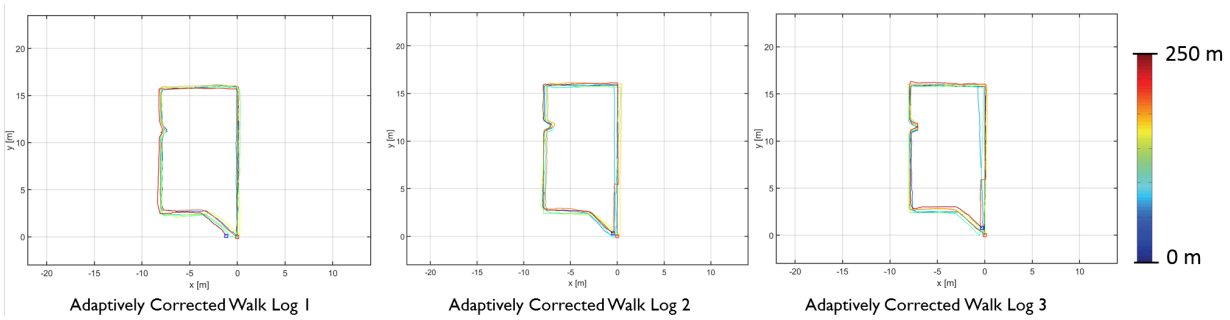


FIGURE 6.10: Adaptively Corrected Forward and Sideways walk trajectories.

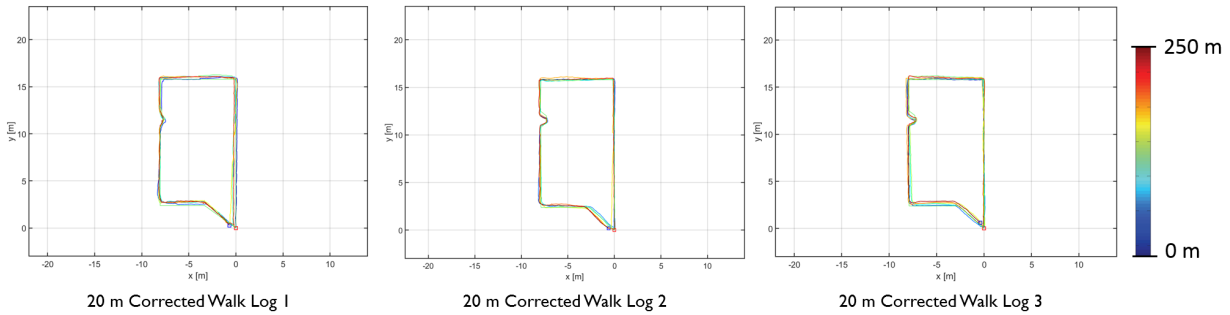


FIGURE 6.11: 20m Corrected Forward and Sideways walk trajectories.

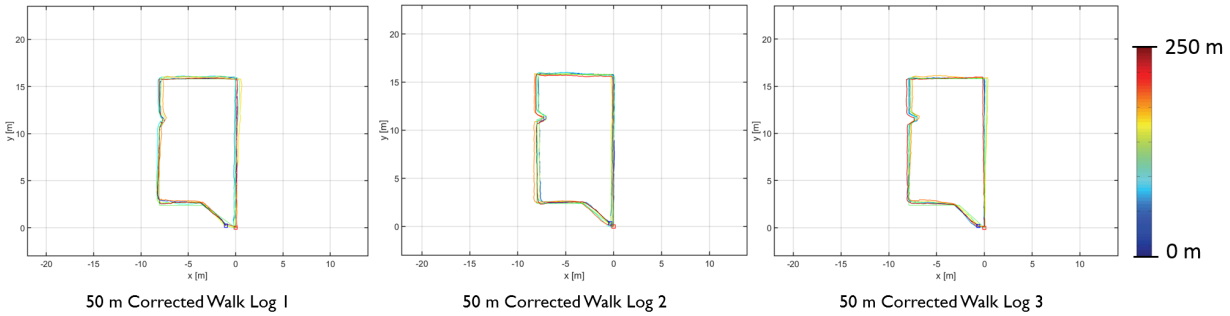


FIGURE 6.12: 50m Corrected Forward and Sideways walk trajectories.

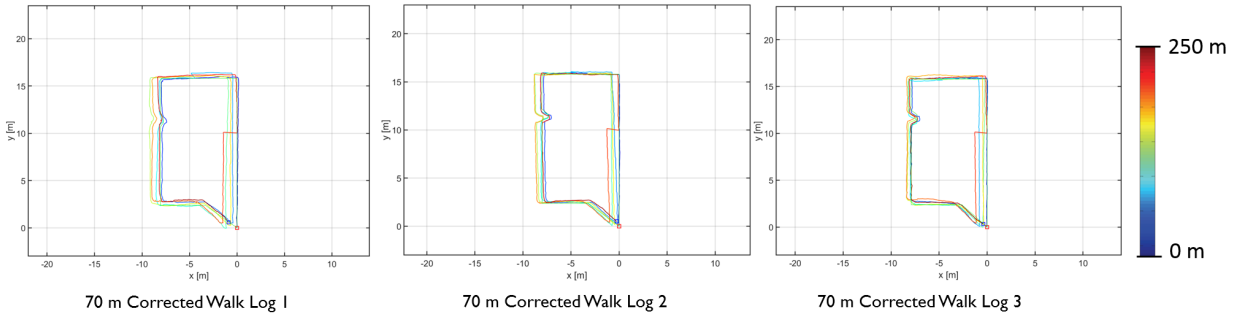


FIGURE 6.13: 70m Corrected Forward and Sideways walk trajectories.

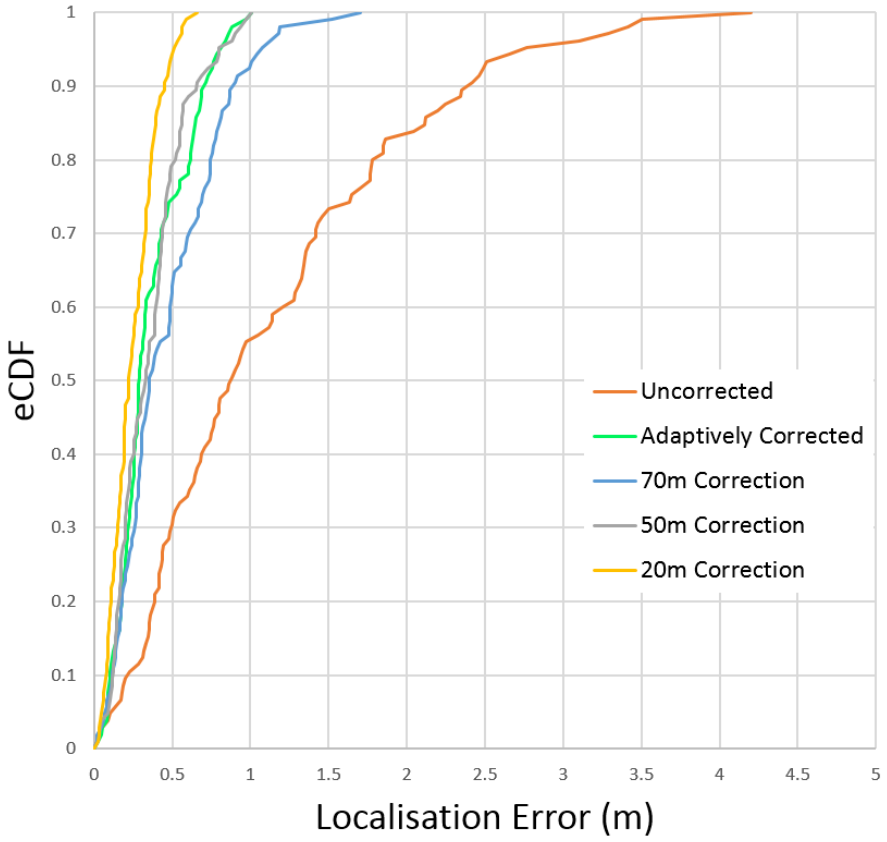


FIGURE 6.14: Combined eCDF values versus localisation errors for the three trajectories.

c) Forward, Backward, and Sideways Walk

The three mixed motion type group had all the motion types being studied. During each loop, the participant walked with the following motion types sequence: 15m forward, then 10m backward, followed by 15m forward, and lastly, 10m sideways. The uncorrected and adaptively corrected walking trajectories are shown in Figure 6.15 and Figure 6.16, respectively. The localisation errors for uncorrected trajectories were 2.25m,

3.94m, and 3.12m for 95% of the walk and for adaptively corrected trajectories were 0.76m, 0.98m, and 0.85m. Similar to the two mixed motion type groups, adaptively corrected trajectories were compared with distance-based correction applied at 20m, 50m, and 70m distance thresholds, and the results are shown in Figure 6.17, Figure 6.18, and Figure 6.19, respectively. The localisation errors versus the eCDF values for the three trajectories are plotted in Figure 6.20.

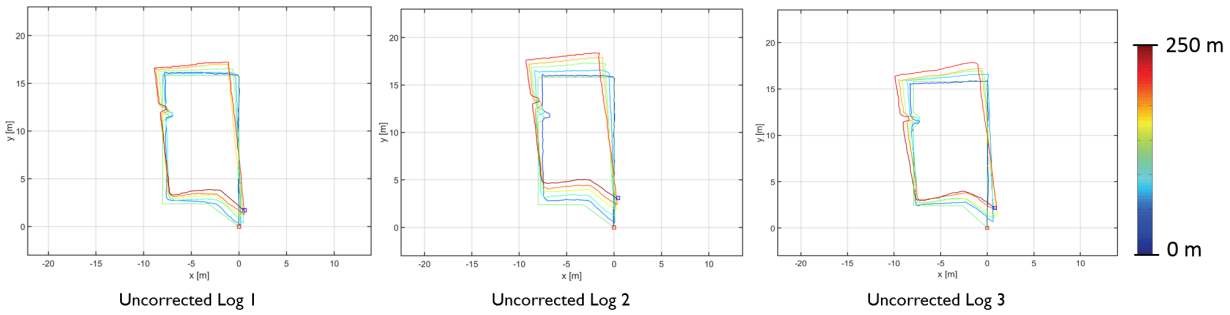


FIGURE 6.15: Uncorrected Forward, Backward, and Sideways walk trajectories.

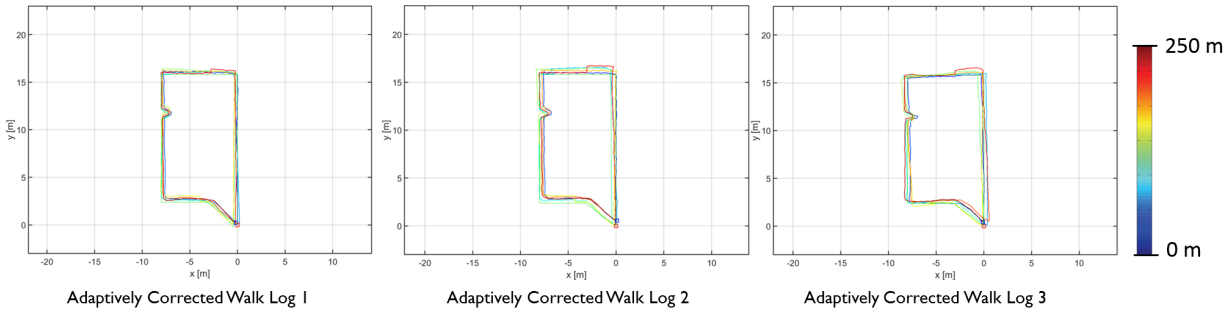


FIGURE 6.16: Adaptively Corrected Forward, Backward, and Sideways walk trajectories.

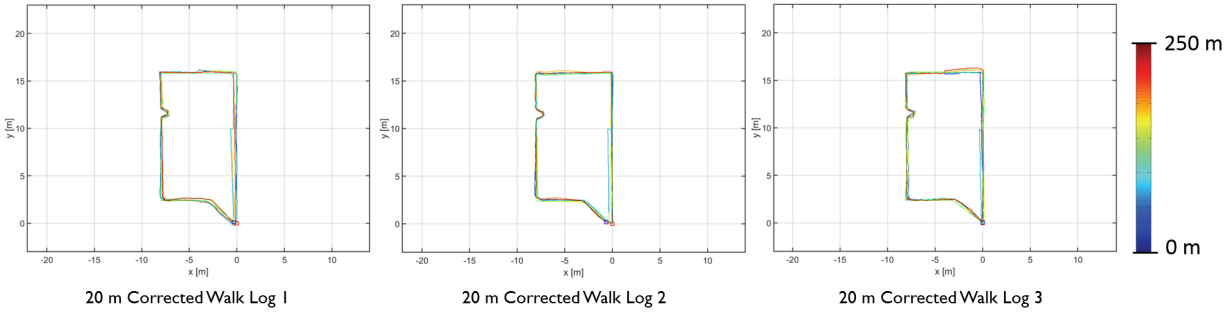


FIGURE 6.17: 20m Corrected Forward, Backward, and Sideways walk trajectories.

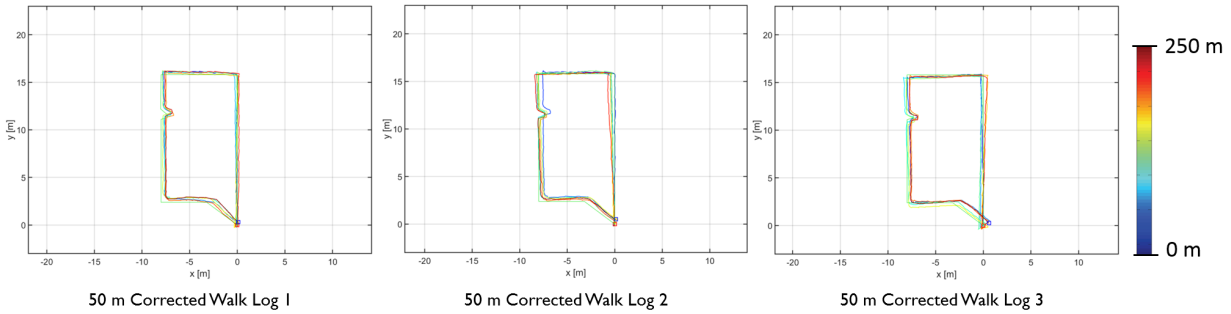


FIGURE 6.18: 50m Corrected Forward, Backward, and Sideways walk trajectories.

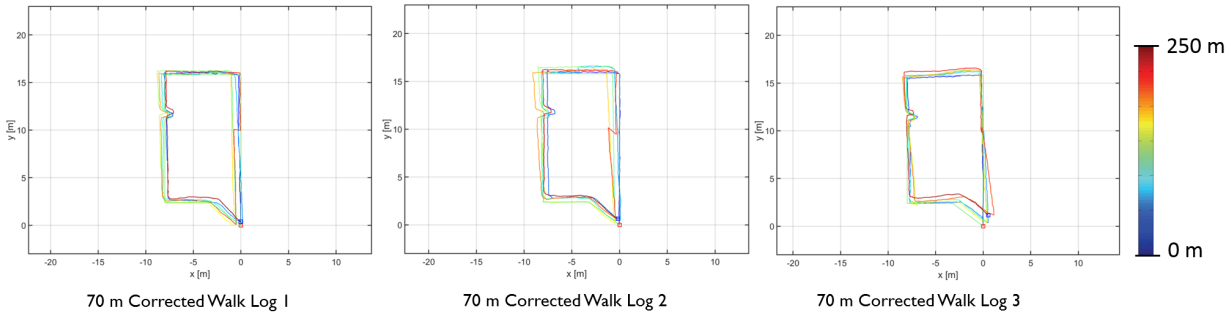


FIGURE 6.19: 70m Corrected Forward, Backward, and Sideways walk trajectories.

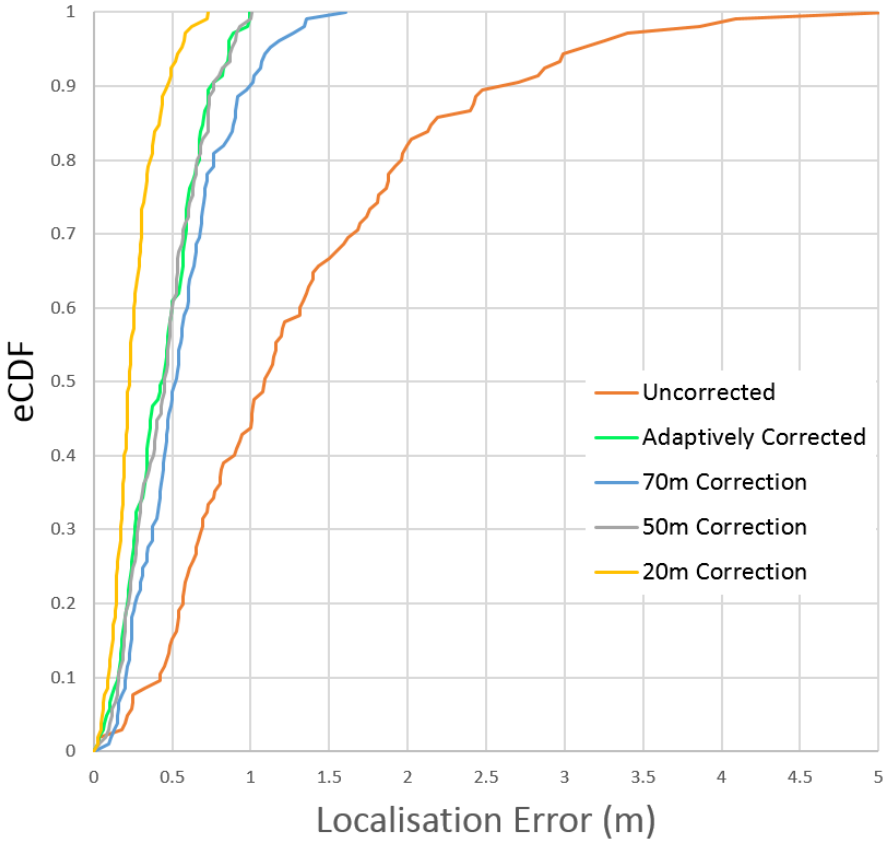


FIGURE 6.20: Combined localisation errors versus eCDF values for the three trajectories.

6.4.2 Discussion

The uncorrected walking trajectories of the forward and backward walking motion types (Figure 6.3) indicate error in heading values which is most likely caused during the $10m$ of the backward walk. The results in Section 5.4.1 also showed deviation in heading values for the backward walk. For the forward and sideways walking motion types, a scaling error parallel to the $10m$ of sideways walk can be seen in Figure 6.9 for all three logs. For the forward, backward, and sideways walking motion types (Figure 6.15), there is an error in scaling as well as heading values due to the presence of both the backwards and sideways motion types. The localisation error at the 95% eCDF are summarised in Table 6.1.

For the two and three mixed motion type trajectories, 60% of the walk consisted of forward walking and 40% of sideways and/or backward motion types. The computed number of corrections by our algorithm for the adaptive algorithm is three for both two and three mixed motion types. The number of corrections required for the distance-based error correction method were: 12 corrections for $20m$ distance threshold, 4 corrections for $50m$ distance threshold, and 2 corrections for $70m$. From these results, Table 6.1 and the eCDF plots in Figures 6.8, 6.14, and 6.20, we can see that the $50m$ distance-based correction has similar accuracy to the adaptive correction approach. This implies 25% reduction in number of corrections required for the adaptive approach in mixed motion types, which saves 1.257 Ws of energy for the $250m$ walked trajectory.

TABLE 6.1: Localisation error at 95% of the walk in mixed motion type trajectories

Type of Correction	**Forward-Backward Walk**	**Forward-Sideways Walk**	**Forward-Backward-Sideways Walk**
Uncorrected	$3.46m$	$2.77m$	$3.12m$
Adaptively Corrected	$0.89m$	$0.80m$	$0.86m$
$20m$ Correction	$0.52m$	$0.52m$	$0.56m$
$50m$ Correction	$0.94m$	$0.80m$	$0.88m$
$70m$ Correction	$1.19m$	$1.08m$	$1.12m$

6.4.3 Limitations

Our results validate our proposition that varying the correction mechanism with the change in motion type reduces the number of corrections required, which in turn saves resources. However, the designed experiment studied the change in the motion type in a

structured way where the participant walked one motion type for a certain distance then switched to the others. In real-life scenarios, there is more variation in motion types, and the transition between them could be more frequent. The accuracy of motion type classification also plays a role. Less accurate classification leads to a wrong error accumulation parameter, resulting in either too frequent corrections or reduced localisation accuracy. As discussed in Section 6.3, classification is still being actively researched. For an assumptive discussion, if we extend the classification accuracy results from Wagstaff et al. (77% for walking, 73% for running, and 78% for climbing stairs) [203] to walking motion types, we will have an average classification accuracy of 75%, which implies if there are five corrections to be applied, two will be wrongly classified and only three will be correct. This reduces the effectiveness of our adaptive approach by 40%. This should be investigated in future research when classification between walking motion types is available. Apart from motion types, participant walking pace and walking surface are the two other factors that will also affect the localisation results as shown by Wahlström, et al. [164]. These two factors should also be considered in future research to make the localisation system more robust.

6.5 Conclusion

In this chapter, we formulated a motion type adaptive correction algorithm to further improve the energy efficiency of the localisation system. Our proposed method is based on the change in error accumulation rate for different motion types. To determine the next correction from the camera system, two inputs were considered: (1) the error accumulation rate of each motion type, and (2) the operational localisation accuracy required by the system. Mixed motion types investigated in this chapter are forward, sideways, and backward walking motion types. Using this approach, our devised localisation system was able to reduce the number of corrections by 25% saving 1.257 Ws of energy for a 250m walked trajectory. **These results validate that modifying the correction mechanism for different motion types reduces the number of corrections while maintaining the positioning accuracy within the required sub-metre range.** Comparison with the distance-based error correction approach also indicates better performance of the motion type adaptive approach.

Chapter 7

Conclusions and Future Work

This thesis has presented a novel indoor localisation system for first responders (pedestrians) in large-scale environments by loosely coupling visual and inertial sensors. This chapter summarises the work done in previous chapter, identifies the main contributions of the thesis, and discusses the potential areas for future research.

7.1 Summary

The motivation for developing an indoor localisation solution and the intended application of the solution were discussed in Chapter 1. In Chapter 2, existing indoor positioning technologies for localising first responders (pedestrians) were surveyed and compared. Chapter 2 also discussed sensor fusion algorithms and the choice of the extended Kalman filter for efficiency and resource conservation to track non-linear states. In Chapter 3, fundamentals of inertial and visual sensors were presented, including the coordinate systems involved, types of errors in inertial sensors and their mitigation, and visual sensor's intrinsic and extrinsic parameter estimation.

Chapter 4 presented the novel loosely coupled sensor fusion approach, which combines the inertial and visual sensors using an extended Kalman filter. The visual sensor was aided by pre-mapped fiducial markers placed in the environment. Accurate mapping of the markers was also explained using indoor surveying equipment. Evaluation over a distance of 250 m walks indicated sub-metre level localisation accuracy.

Chapter 5 described the proposed distance-based error correction method, which used a distance-based error correction approach to improve the energy efficiency of the system. The method to accurately measure the power consumption of each device in the localisation system was also explained. The evaluation of the system was extended to two other walking motion types (sideways and backward walks) which have not been studied for longer distances in the past. The study results showed a decrease in localisation accuracy for both sideways and backward motion types. Results also revealed a 70% decrease in energy consumption with the distance-based error correction approach.

Chapter 6 further studied the effect of motion types with the focus on making the system more adaptive in a mixed-motion type walking scenario, which is closer to a real-life scenario when walking varies between motion types. A novel motion type adaptive correction method was devised to determine the next correction from the visual sensor using motion type as an input. The evaluation of the adaptive correction approach indicated a 25% decrease in the number of corrections, which further reduced the power consumption of the system for long-term tracking.

7.2 Contributions

The first main contribution of this thesis is the development and evaluation of the loosely coupled sensor fusion system by combining inertial and visual sensors with an extended Kalman filter. The developed localisation system used off-the-shelf inertial and visual sensors and showed that an accurate long-term position and orientation tracking is possible with consumer-grade medium-cost sensors. The localisation system is also scalable to a larger area by the addition of new fiducial markers in the environment. As there is no power source required to operate the fiducial markers, this makes them ideal for applications which have limited or no power source available in the environment (refer to Sections 1.1 and 1.2 for details).

The second main contribution of this thesis is the distance-based error correction method that reduces the localisation system's energy consumption by 70%. This is an important contribution as the current highest energy density of electric batteries (Lithium-ion, 129 Wh/kg [233]) is limited, and to localise pedestrians for longer time intervals more battery weight has to be added. For example, the developed visual-inertial sensor fusion

system would last approximately 24 hours on a 65 Wh battery which weighs about 500g. However, with the distance-based error correction, it can last for about 85 hours. More weight is not viable for pedestrian localisation or other applications requiring movement. The devised system is a step towards a long-duration localisation system; however, more work needs to be done to make the localisation solution real-world deployable. Section 7.4 discusses some of the possible future directions.

The third main contribution of this thesis is the formulation of a motion type adaptive model which changes the error correction mechanism when the motion type changes. This reduces the number of corrections required to maintain sub-metre localisation accuracy by 25%. Reduction in the number of corrections implies less battery usage which in turn translates into longer battery life or less battery weight.

Apart from these main contributions, this thesis has also identified unique acceleration and angular velocity signals for forward, backward, and sideways walking motion types (Section 5.4.2). These signals could be useful in segmenting the motion types in future research. The effect of these three motion types on localisation accuracy was also studied over a longer distance (250 m), which has not been done in the past research.

This thesis also provides documentation on accurately measuring the power consumption of inertial and visual sensors. This is important as highlighted in Section 5.3, as a higher burden voltage could lead to a significantly higher current measurement, which has happened in past research [25]. Section 5.3 is a valuable reference for future researchers working on improving the energy efficiency of such systems, not only in the localisation space but also in other fields which involve low-power and high dynamic range current measurements. Accurate fiducial marker mapping for large-scale indoor environments using consumer-grade surveying equipment is another process that has not been clearly documented in the past literature to the best of our knowledge. Section 4.2.1 could be used as a reference in the future for accurate fiducial marker mapping.

7.3 Limitations

To make the proposed localisation solution generalisable and applicable to real-world first responder application, the following limitations should be addressed in future research:

Testing on Multiple Users

Most of the experimental evaluation was performed on a single test subject due to the COVID pandemic. To generalise the results, the localisation system should be tested with a diverse group of participants covering different age ranges, heights, and gender. The proposed localisation algorithm does not use any physical parameters of the user, so the findings are transferable to users with normal gait; however, further evaluation is required for conclusive results.

Walking Surface and Variation in Test Environment

The localisation system was mainly evaluated in an office environment with carpet as the walking surface. From Wahlström et al.'s work [15] we know that the walking surface could affect the localisation accuracy; therefore, variation of the walking surfaces, including both hard surfaces (concrete, asphalt, gravel road etc.), and soft (dirt, grass and sand road etc.) should be tested. Also, testing at the urban training facilities shown in Figures 1.1, 1.2, and 1.3 would offer insights into the performance of the localisation accuracy in the real-world application.

Motion Types and Transition between them

To keep the localisation solution more tractable, only the walking motion type has been focused on in this research. Aside from walking, the actual application might involve other motion types such as running, climbing, jogging, crawling, standing, jumping, shuffling, crouching, crouch walking, and knee-dragging [14, 15, 199–201]. There has been some research done on the walking, running, and climbing motion types, but a lot less on the others. The remaining motion types and transition between them should be further investigated to understand their effect on the localisation accuracy. This point has been discussed in more detail in Chapter 6, which could be referred to in future research.

Ground Truth Method

The ground truth method used for to evaluate our system was an offline manual process that took a lot of time of video analysis, and it should be automated. As discussed in Section 4.4.2, motion capturing systems are too expensive to install in large-scale indoor environments and do not work in spaces with a lot of sunlight. A potential direction to automate the ground truth method is to use a drone that is wirelessly connected and temporally aligned with the laptop. This has not yet been explored in past research, and if created, will offer a portable solution that can be used in different indoor environments. A good candidate drone to create this method is Tello drone [234], that comes with an opensource SDK and forward-facing camera which could be used to determine pedestrian location. The drone could be modified with a ceiling facing camera that would have fiducial markers in the field of view to accurately track the drone's position in the global frame of reference. An automated low-cost and accurate ground truth method would be an important contribution in the indoor positioning space as identified in past research [235].

Data Recording and Processing Hardware

For the proposed localisation system, the data from both visual and inertial sensors were recorded on a laptop, which is bulky and not feasible to carry for all motion types. The laptop (Razer Blade 13.3 inch) used for the proposed system weighed less than 1.3 kg, which was manageable while walking; however, it would hinder movement during running, jumping, or other more dynamic motion types. The laptop should be replaced with a Raspberry Pi for other motion types. Since the application is after-action review (non-real-time), the sensor data from the pedestrian does not need to be transmitted back to a central computer, but could be stored with the timestamps and processed afterwards.

7.4 Future Research

Apart from the limitations identified above, there are also other avenues that could be explored in the future. Some of these are concerned with improving energy efficiency, while others address the localisation.

Areas to Further Increase Energy Efficiency

The localisation system for this thesis used off-the-shelf visual and inertial sensors. Creating customised hardware solely focused on energy efficiency using low-cost electronic components has not been investigated, and it could further increase energy efficiency.

Multiple users, also referred to as cooperative localisation, is another research area that has not been studied from an energy efficiency point-of-view [99, 236, 237]. Using a position and heading direction estimate of one person in a group, determining other members' localisation in reference to the accurately tracked person, and managing the energy efficiency of the group as a whole could lead to a further improvement in energy efficiency.

Map-matching methods use walls and other physical constraints in the environment to increase the accuracy of a PDR-based localisation system [2, 180, 238]. Using map-matching constraints to optimise the correction mechanism for energy efficiency is another research topic that could be explored.

Motion Types

As discussed in Section 6.3, forward, backward, and sideways walking motion types have subtle variation in acceleration and angular velocity signals compared to walking, running, and crawling motion types which have distinguishing features. There has not been any research focused on segmenting these motion types. Signals found in this thesis could be a source of accurate segmentation for these motion types.

The adaptive approach described in Chapter 6 had three walking motion types as an input. The devised adaptive formulation could be extended to include other motion types involved in first responder training exercises, such as crawling, crouching, and climbing.

Invisible Fiducial Markers

To provide localisation in low-visibility environments for training environments, such as dark or smoky conditions, fiducial markers can be printed with ultraviolet (UV) or infrared (IR) ink and detected with a UV or an IR camera. Our research in this area indicated that a normal visible light spectrum camera can be converted to a UV or IR camera by removing the reflection filters [239]. There has been limited research in this domain [240, 241], and it has potential to extend our localisation solutions for these environments. Aesthetics is another reason why fiducial markers are not used in applications such as localisation in large grocery stores and shopping malls [242, 243]. With invisible markers, the visual appearance of the indoor spaces would have less effect, and it should be further investigated.

Creating buildings that are friendly to visually-impaired individuals is another application that could be explored with invisible markers. As the aesthetic impact of fiducial markers could prevent their installation, and marker-less SLAM systems (discussed in Section 2.1.1) have higher chance of navigation failure with the repetitive featureless pattern in indoor spaces, buildings could have invisible markers installed at regular intervals on the ceiling or doors similar to Braille signage, to provide navigation for the visually impaired.

Bibliography

[1] V. Dimitrov, J. Vazquez, and T. Padır, "Locater: Localization and accountability technologies for emergency responders," in *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*. IEEE, 2017, pp. 1–4.

[2] A. F. G. G. Ferreira, D. M. A. Fernandes, A. P. Catarino, and J. L. Monteiro, "Localization and positioning systems for emergency responders: A survey," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 4, pp. 2836–2870, 2017.

[3] F. Zafari, A. Gkelias, and K. K. Leung, "A survey of indoor localization systems and technologies," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2568–2599, 2019.

[4] R. Mautz, "Indoor positioning technologies," 2012.

[5] J. Rawlings, "The fake city streets where cops learn riot control," July 05, 2014 [Online]. [Online]. Available: https://www.wired.com/2014/05/james-rawlings-architecture-of-conflict/

[6] R. Wheeler, "Machines for training: Military performance and embodied knowledge," June 12, 2015 [Online]. [Online]. Available: http://www.incendiarytraces.org/articles/2015/9/11/machines-for-training-military-performance-and-embodied-knowledge

[7] A. Alarifi, A. Al-Salman, M. Alsaleh, A. Alnafessah, S. Al-Hadhrami, M. A. Al-Ammar, and H. S. Al-Khalifa, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors*, vol. 16, no. 5, p. 707, 2016.

[8] F. Seco and A. R. Jimenez, "Autocalibration of a wireless positioning network with a fastslam algorithm," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2017, pp. 1–8.

[9] H. Park, J. Noh, and S. Cho, "Three-dimensional positioning system using bluetooth low-energy beacons," *International Journal of Distributed Sensor Networks*, vol. 12, no. 10, p. 1550147716671720, 2016.

[10] N. Joubert, T. G. Reid, and F. Noble, "Developments in modern gnss and its impact on autonomous vehicle architectures," in *2020 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2020, pp. 2029–2036.

[11] G. Li, E. Geng, Z. Ye, Y. Xu, J. Lin, and Y. Pang, "Indoor positioning algorithm based on the improved rssi distance model," *Sensors*, vol. 18, no. 9, p. 2820, 2018.

[12] E. Foxlin, "Pedestrian tracking with shoe-mounted inertial sensors," *IEEE Computer Graphics and Applications*, vol. 25, no. 6, pp. 38–46, 2005.

[13] I. Skog, P. Handel, J.-O. Nilsson, and J. Rantakokko, "Zero-velocity detection—an algorithm evaluation," *IEEE transactions on biomedical engineering*, vol. 57, no. 11, pp. 2657–2666, 2010.

[14] B. Wagstaff, V. Peretroukhin, and J. Kelly, "Improving foot-mounted inertial navigation through real-time motion classification," *arXiv preprint arXiv:1707.01152*, 2017.

[15] J. Wahlström and I. Skog, "Fifteen years of progress at zero velocity: A review," *IEEE Sensors Journal*, vol. 21, no. 2, pp. 1139–1151, 2020.

[16] R. Mur-Artal and J. D. Tardós, "Orb-slam2: An open-source slam system for monocular, stereo, and rgb-d cameras," *IEEE Transactions on Robotics*, vol. 33, no. 5, pp. 1255–1262, 2017.

[17] C. Cadena, L. Carlone, H. Carrillo, Y. Latif, D. Scaramuzza, J. Neira, I. Reid, and J. J. Leonard, "Past, present, and future of simultaneous localization and mapping: Toward the robust-perception age," *IEEE Transactions on robotics*, vol. 32, no. 6, pp. 1309–1332, 2016.

[18] B. Bodin, H. Wagstaff, S. Saecdi, L. Nardi, E. Vespa, J. Mawer, A. Nisbet, M. Luján, S. Furber, A. J. Davison *et al.*, "Slambench2: Multi-objective head-to-head benchmarking for visual slam," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 3637–3644.

[19] J. Fuentes-Pacheco, J. Ruiz-Ascencio, and J. M. Rendón-Mancha, "Visual simultaneous localization and mapping: a survey," *Artificial intelligence review*, vol. 43, no. 1, pp. 55–81, 2015.

[20] H. Taira, M. Okutomi, T. Sattler, M. Cimpoi, M. Pollefeys, J. Sivic, T. Pajdla, and A. Torii, "Inloc: Indoor visual localization with dense matching and view synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 7199–7209.

[21] H. Taira, I. Rocco, J. Sedlar, M. Okutomi, J. Sivic, T. Pajdla, T. Sattler, and A. Torii, "Is this the right place? geometric-semantic pose verification for indoor visual localization," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4373–4383.

[22] X. Zhang, J. Lin, Q. Li, T. Liu, and Z. Fang, "Continuous indoor visual localization using a spatial model and constraint," *IEEE Access*, vol. 8, pp. 69 800–69 815, 2020.

[23] X. Chen, K. W. Nixon, and Y. Chen, "Practical power consumption analysis with current smartphones," in *2016 29th IEEE International System-on-Chip Conference (SOCC)*. IEEE, 2016, pp. 333–337.

[24] X. Chen, Y. Chen, Z. Ma, and F. C. Fernandes, "How is energy consumed in smartphone display applications?" in *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, 2013, pp. 1–6.

[25] S. Zhu, H. Anderson, and Y. Wang, "Reducing the power consumption of an imu-based gait measurement system," in *Pacific-Rim Conference on Multimedia*. Springer, 2012, pp. 105–116.

[26] *BNO055 Datasheet, Document BST-BNO055-DS000-14*, Bosch, 06 2016, revision: 1.4.

[27] *ICM-20948 Datasheet, Document DS-000189*, TDK InvenSense, 02 2017, revision: 1.3.

[28] *LSM9DS1 Datasheet, Document DocID025715*, STMicroelectronics, 03 2015, rev 3.

[29] *MTi 1-series Datasheet, Document MT0512P*, XSens Technologies, 12 2019, rev.2019.A.

[30] G. Fusco and J. M. Coughlan, "Indoor localization using computer vision and visual-inertial odometry," in *International Conference on Computers Helping People with Special Needs*. Springer, 2018, pp. 86–93.

[31] M. Li and A. I. Mourikis, "High-precision, consistent ekf-based visual-inertial odometry," *The International Journal of Robotics Research*, vol. 32, no. 6, pp. 690–711, 2013.

[32] A. Poulose and D. S. Han, "Hybrid indoor localization using imu sensors and smartphone camera," *Sensors*, vol. 19, no. 23, p. 5084, 2019.

[33] H. Khan, A. Clark, G. Woodward, and R. W. Lindeman, "Improved position accuracy of foot-mounted inertial sensor by discrete corrections from vision-based fiducial marker tracking," *Sensors*, vol. 20, no. 18, p. 5031, 2020.

[34] R. M. Clifford, H. Khan, S. Hoermann, M. Billinghurst, and R. W. Lindeman, "Development of a multi-sensory virtual reality training simulator for airborne firefighters supervising aerial wildfire suppression," in *2018 IEEE Workshop on Augmented and Virtual Realities for Good (VAR4Good)*. IEEE, 2018, pp. 1–5.

[35] R. M. Clifford, H. Khan, M. Billinghurst, and R. W. Lindeman, "The effect of immersive displays on situation awareness in virtual environments for aerial firefighting air attack supervisor training," in *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, 2018, pp. 1–2.

[36] D. Barry, M. Shah, M. Keijsers, H. Khan, and B. Hopman, "Xyolo: A model for real-time object detection in humanoid soccer on low-end hardware," in *2019 International Conference on Image and Vision Computing New Zealand (IVCNZ)*. IEEE, 2019, pp. 1–6.

[37] J. Kunhoth, A. Karkar, S. Al-Maadeed, and A. Al-Ali, "Indoor positioning and wayfinding systems: a survey," *Human-centric Computing and Information Sciences*, vol. 10, pp. 1–41, 2020.

[38] G. Welch and E. Foxlin, "Motion tracking: No silver bullet, but a respectable arsenal," *IEEE Computer graphics and Applications*, vol. 22, no. 6, pp. 24–38, 2002.

[39] R. Muñoz-Salinas, M. J. Marín-Jimenez, E. Yeguas-Bolivar, and R. Medina-Carnicer, "Mapping and localization from planar markers," *Pattern Recognition*, vol. 73, pp. 158–171, 2018.

[40] G. Huang, "Visual-inertial navigation: A concise review," in *2019 international conference on robotics and automation (ICRA)*. IEEE, 2019, pp. 9572–9582.

[41] T. Taketomi, H. Uchiyama, and S. Ikeda, "Visual slam algorithms: a survey from 2010 to 2016," *IPSJ Transactions on Computer Vision and Applications*, vol. 9, no. 1, pp. 1–11, 2017.

[42] L. Nardi, B. Bodin, M. Z. Zia, J. Mawer, A. Nisbet, P. H. J. Kelly, A. J. Davison, M. Luján, M. F. P. O'Boyle, G. Riley, N. Topham, and S. Furber, "Introducing slambench, a performance and accuracy benchmarking methodology for slam," in *IEEE Intl. Conf. on Robotics and Automation (ICRA)*, May 2015, arXiv:1410.2167.

[43] E. Brachmann and C. Rother, "Learning less is more-6d camera localization via 3d surface regression," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4654–4662.

[44] A. Kendall and R. Cipolla, "Geometric loss functions for camera pose regression with deep learning," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 5974–5983.

[45] M. Fiala, "Comparing artag and artoolkit plus fiducial marker systems," in *Haptic Audio Visual Environments and their Applications, 2005. IEEE International Workshop on*. IEEE, 2005, pp. 6–pp.

[46] G. Goronzy, M. Pelka, and H. Hellbrück, "Qrpos: Indoor positioning system for self-balancing robots based on qr codes," in *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2016, pp. 1–8.

[47] S.-J. Lee, G. Tewolde, J. Lim, and J. Kwon, "Qr-code based localization for indoor mobile robot with validation using a 3d optical tracking instrument," in *2015 IEEE International Conference on Advanced Intelligent Mechatronics (AIM)*. IEEE, 2015, pp. 965–970.

[48] B. Marques, R. Carvalho, P. Dias, M. Oliveira, C. Ferreira, and B. S. Santos, "Evaluating and enhancing google tango localization in indoor environments using fiducial markers," in *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, 2018, pp. 142–147.

[49] O. J. Woodman, "An introduction to inertial navigation," University of Cambridge, Computer Laboratory, Tech. Rep., 2007.

[50] L. Ojeda and J. Borenstein, "Non-gps navigation with the personal dead-reckoning system," in *Unmanned Systems Technology IX*, vol. 6561. International Society for Optics and Photonics, 2007, p. 65610C.

[51] R. Harle, "A survey of indoor inertial positioning systems for pedestrians." *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 1281–1293, 2013.

[52] R. G. Stirling, "Development of a pedestrian navigation system using shoe mounted sensors," Ph.D. dissertation, University of Alberta, 2004.

[53] Q. Ladetto, J. Van Seeters, S. Sokolowski, Z. Sagan, and B. Merminod, "Digital magnetic compass and gyroscope for dismounted soldier position and navigation," in *NATO-RTO meetings*, 2002.

[54] A. R. Jimenez, F. Seco, C. Prieto, and J. Guevara, "A comparison of pedestrian dead-reckoning algorithms using a low-cost mems imu," in *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*. IEEE, 2009, pp. 37–42.

[55] C. Fischer, P. T. Sukumar, and M. Hazas, "Tutorial: Implementing a pedestrian tracker using inertial sensors," *IEEE pervasive computing*, vol. 12, no. 2, pp. 17–27, 2012.

[56] J. Callmer, D. Törnqvist, and F. Gustafsson, "Probabilistic stand still detection using foot mounted imu," in *Information Fusion (FUSION), 2010 13th Conference on*. IEEE, 2010, pp. 1–7.

[57] R. F. Alonso, E. Z. Casanova, and J. G. García-Bermejo, "Pedestrian tracking using inertial sensors," *Journal of Physical Agents*, vol. 3, no. 1, pp. 35–43, 2009.

[58] A. R. Jiménez, "Tutorial: Pedestrian dead-reckoning," 2017.

[59] B. Wang, X. Liu, B. Yu, R. Jia, and X. Gan, "Pedestrian dead reckoning based on motion mode recognition using a smartphone," *Sensors*, vol. 18, no. 6, p. 1811, 2018.

[60] H. Weinberg, "Using the adxl202 in pedometer and personal navigation applications," *Analog Devices AN-602 application note*, vol. 2, no. 2, pp. 1–6, 2002.

[61] M. Vezočnik and M. B. Juric, "Average step length estimation models' evaluation using inertial sensors: a review," *IEEE Sensors Journal*, vol. 19, no. 2, pp. 396–403, 2018.

[62] S. Godha, G. Lachapelle, and M. E. Cannon, "Integrated gps/ins system for pedestrian navigation in a signal degraded environment," in *In Proceedings of the 19th International Technical Meetings of the Satellite Division of the Institute of Navigation, Forth*, 2006, pp. 2151–2164.

[63] F. Cavallo, A. M. Sabatini, and V. Genovese, "A step toward gps/ins personal navigation systems: real-time assessment of gait by foot inertial sensing," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Aug 2005, pp. 1187–1191.

[64] Y. S. Suh and S. Park, "Pedestrian inertial navigation with gait phase detection assisted zero velocity updating," in *Autonomous Robots and Agents, 2009. ICARA 2009. 4th International Conference on.* IEEE, 2009, pp. 336–341.

[65] C. Toth, D. A. Grejner-Brzezinska, and S. Moafipoor, "Pedestrian tracking and navigation using neural networks and fuzzy logic," in *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on.* IEEE, 2007, pp. 1–6.

[66] Q. Wang, X. Zhang, X. Chen, R. Chen, W. Chen, and Y. Chen, "A novel pedestrian dead reckoning algorithm using wearable emg sensors to measure walking strides," in *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, Oct 2010.

[67] A. Kotanen, M. Hannikainen, H. Leppakoski, and T. Hamalainen, "Positioning with ieee 802.11 b wireless lan," in *14th IEEE Proceedings on Personal, Indoor and Mobile Radio Communications, 2003. PIMRC 2003.*, vol. 3. IEEE, 2003, pp. 2218–2222.

[68] G. Chen, X. Meng, Y. Wang, Y. Zhang, P. Tian, and H. Yang, "Integrated wifi/p-dr/smartphone using an unscented kalman filter algorithm for 3d indoor localization," *Sensors*, vol. 15, no. 9, pp. 24 595–24 614, 2015.

[69] T. T. Khanh, V. Nguyen, X.-Q. Pham, and E.-N. Huh, "Wi-fi indoor positioning and navigation: a cloudlet-based cloud computing approach," *Human-centric Computing and Information Sciences*, vol. 10, no. 1, pp. 1–26, 2020.

[70] S. Xia, Y. Liu, G. Yuan, M. Zhu, and Z. Wang, "Indoor fingerprint positioning based on wi-fi: An overview," *ISPRS International Journal of Geo-Information*, vol. 6, no. 5, p. 135, 2017.

[71] V. Moghtadaiee and A. G. Dempster, "Design protocol and performance analysis of indoor fingerprinting positioning systems," *Physical Communication*, vol. 13, pp. 17–30, 2014.

[72] A. R. J. Ruiz, F. S. Granja, J. C. P. Honorato, and J. I. G. Rosas, "Pedestrian indoor navigation by aiding a foot-mounted imu with rfid signal strength measurements," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*. IEEE, 2010, pp. 1–7.

[73] J. Peng, M. Zhu, and K. Zhang, "New algorithms based on sigma point kalman filter technique for multi-sensor integrated rfid indoor/outdoor positioning," in *2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN) Guimaraes, Portugal*, 2011, pp. 21–23.

[74] K. Dziadak, B. Kumar, and J. Sommerville, "Rfid in the built environment: Buried asset locating systems," in *Intelligent Computing in Engineering and Architecture*. Springer, 2006, pp. 153–162.

[75] M. Fujimoto, E. Nakamori, A. Inada, Y. Oda, T. Wada, K. Mutsuura, and H. Okada, "A broad-typed multi-sensing-range method for indoor position estimation of passive rfid tags," in *Proceedings of the 2011 International Conference on Indoor Positioning and Indoor Navigation (IPIN'2011)*, 2011.

[76] S. Y. Seidel and T. S. Rappaport, "914 mhz path loss prediction models for indoor wireless communications in multifloored buildings," *IEEE transactions on Antennas and Propagation*, vol. 40, no. 2, pp. 207–217, 1992.

[77] A. D. Koutsou, F. Seco, A. R. Jiménez, J. O. Roa, J. L. Ealo, C. Prieto, and J. Guevara, "Preliminary localization results with an rfid based indoor guiding system," in *Intelligent Signal Processing, 2007. WISP 2007. IEEE International Symposium on.* IEEE, 2007, pp. 1–6.

[78] F. Seco, C. Plagemann, A. R. Jiménez, and W. Burgard, "Improving rfid-based indoor positioning accuracy using gaussian processes," in *Indoor Positioning and Indoor Navigation (IPIN).* IEEE, 2010, pp. 1–8.

[79] A. Cazzorla, G. De Angelis, A. Moschitta, M. Dionigi, F. Alimenti, and P. Carbone, "A 5.6-ghz uwb position measurement system," *IEEE Transactions on Instrumentation and Measurement*, vol. 62, no. 3, pp. 675–683, 2012.

[80] A. R. J. Ruiz and F. S. Granja, "Comparing ubisense, bespoon, and decawave uwb location systems: Indoor performance analysis," *IEEE Transactions on instrumentation and Measurement*, vol. 66, no. 8, pp. 2106–2117, 2017.

[81] R. Heydon and N. Hunn, "Bluetooth low energy," *CSR Presentation, Bluetooth SIG https://www. bluetooth. org/DocMan/handlers/DownloadDoc. ashx*, 2012.

[82] B. Jallow and W.-T. Hong, "Evaluation of indoor positioning based on ibeacon and pi-beacon," in *2018 International Conference on System Science and Engineering (ICSSE).* IEEE, 2018, pp. 1–4.

[83] A. Dasgupta, R. Nagaraj, and K. Nagamani, "An internet of things platform with google eddystone beacons," *Journal of Software Engineering and Applications*, vol. 9, no. 6, pp. 291–295, 2016.

[84] L. Yang, B. Li, H. Li, and Y. Shen, "ibeacon/wifi signal characteristics analysis for indoor positioning using mobile phone," in *China Satellite Navigation Conference.* Springer, 2017, pp. 405–416.

[85] Y. Li, M. Scanavino, E. Capello, F. Dabbene, G. Guglieri, and A. Vilardi, "A novel distributed architecture for uav indoor navigation," *Transportation research procedia*, vol. 35, pp. 13–22, 2018.

[86] A. Harter, A. Hopper, P. Steggles, A. Ward, and P. Webster, "The anatomy of a context-aware application," *Wireless Networks*, vol. 8, no. 2, pp. 187–197, 2002.

[87] N. B. Priyantha, A. Chakraborty, and H. Balakrishnan, "The cricket location-support system," in *Proceedings of the 6th annual international conference on Mobile computing and networking*. ACM, 2000, pp. 32–43.

[88] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 447–458.

[89] M. Maheepala, A. Z. Kouzani, and M. A. Joordens, "Light-based indoor positioning systems: A review," *IEEE Sensors Journal*, vol. 20, no. 8, pp. 3971–3995, 2020.

[90] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, "Epsilon: A visible light based positioning system," in *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, ser. NSDI'14. USA: USENIX Association, 2014, p. 331–343.

[91] S. Bastiaens, "Visible light positioning using an aperture based receiver," *Master Thesis*, 2017.

[92] Y. Bai, W. Jia, H. Zhang, Z.-H. Mao, and M. Sun, "Helping the blind to find the floor of destination in multistory buildings using a barometer," in *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*. IEEE, 2013, pp. 4738–4741.

[93] H. Xia, X. Wang, Y. Qiao, J. Jian, and Y. Chang, "Using multiple barometers to detect the floor location of smart phones with built-in barometric sensors for indoor positioning," *Sensors*, vol. 15, no. 4, pp. 7857–7877, 2015.

[94] Z. Xu, J. Wei, J. Zhu, and W. Yang, "A robust floor localization method using inertial and barometer measurements," in *Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2017, pp. 1–8.

[95] S.-S. Kim, J.-W. Kim, and D.-S. Han, "Floor detection using a barometer sensor in a smartphone," in *Indoor Positioning and Indoor Navigation (IPIN)*, 2017, pp. 1–9.

[96] A. Rai, K. K. Chintalapudi, V. N. Padmanabhan, and R. Sen, "Zee: Zero-effort crowdsourcing for indoor localization," in *Proceedings of the 18th annual international conference on Mobile computing and networking*. ACM, 2012, pp. 293–304.

[97] M. H. Afzal, V. Renaudin, and G. Lachapelle, "Assessment of indoor magnetic field anomalies using multiple magnetometers," in *ION GNSS*, vol. 10, 2010, pp. 21–24.

[98] D. Roetenberg, P. J. Slycke, and P. H. Veltink, "Ambulatory position and orientation tracking fusing magnetic and inertial sensing," *IEEE Transactions on Biomedical Engineering*, vol. 54, no. 5, pp. 883–890, 2007.

[99] F. Seco and A. R. Jiménez, "Smartphone-based cooperative indoor localization with rfid technology," *Sensors*, vol. 18, no. 1, p. 266, 2018.

[100] M. J. Caruso, "Applications of magnetic sensors for low cost compass systems," in *Position Location and Navigation Symposium*. IEEE, 2000, pp. 177–184.

[101] A. Yassin, Y. Nasser, M. Awad, A. Al-Dubai, R. Liu, C. Yuen, R. Raulefs, and E. Aboutanios, "Recent advances in indoor localization: A survey on theoretical approaches and applications," *IEEE Communications Surveys & Tutorials*, vol. 19, no. 2, pp. 1327–1346, 2016.

[102] G. Bishop, G. Welch *et al.*, "An introduction to the kalman filter," *Proc of SIGGRAPH, Course*, vol. 8, no. 27599-3175, p. 59, 2001.

[103] D. Fox, S. Thrun, and W. Burgard, "Probabilistic robotics," 2005.

[104] A. Doucet, N. De Freitas, and N. Gordon, "An introduction to sequential monte carlo methods," in *Sequential Monte Carlo methods in practice*. Springer, 2001, pp. 3–14.

[105] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Transactions on signal processing*, vol. 50, no. 2, pp. 174–188, 2002.

[106] B. Zhou and X. Liao, "Particle filter and levy flight-based decomposed multi-objective evolution hybridized particle swarm for flexible job shop greening scheduling with crane transportation," *Applied Soft Computing*, vol. 91, p. 106217, 2020.

[107] Z. Chen, Q. Zhu, and Y. C. Soh, "Smartphone inertial sensor-based indoor localization and tracking with ibeacon corrections," *IEEE Transactions on Industrial Informatics*, vol. 12, no. 4, pp. 1540–1549, 2016.

[108] A. Correa, M. Barcelo, A. Morell, and J. L. Vicario, "A review of pedestrian indoor positioning systems for mass market applications," *Sensors*, vol. 17, no. 8, p. 1927, 2017.

[109] Y. Wu, H.-B. Zhu, Q.-X. Du, and S.-M. Tang, "A survey of the research status of pedestrian dead reckoning systems based on inertial sensors," *International Journal of Automation and Computing*, vol. 16, no. 1, pp. 65–83, 2019.

[110] A. I. Mourikis and S. I. Roumeliotis, "A multi-state constraint kalman filter for vision-aided inertial navigation," in *Proceedings 2007 IEEE International Conference on Robotics and Automation*. IEEE, 2007, pp. 3565–3572.

[111] S. Weiss, M. W. Achtelik, S. Lynen, M. Chli, and R. Siegwart, "Real-time onboard visual-inertial state estimation and self-calibration of mavs in unknown environments," in *2012 IEEE international conference on robotics and automation*. IEEE, 2012, pp. 957–964.

[112] J. A. Hesch, D. G. Kottas, S. L. Bowman, and S. I. Roumeliotis, "Consistency analysis and improvement of vision-aided inertial navigation," *IEEE Transactions on Robotics*, vol. 30, no. 1, pp. 158–176, 2013.

[113] Z. Huai and G. Huang, "Robocentric visual-inertial odometry," in *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2018, pp. 6319–6326.

[114] N. Trawny, A. I. Mourikis, S. I. Roumeliotis, A. E. Johnson, and J. F. Montgomery, "Vision-aided inertial navigation for pin-point landing using observations of mapped landmarks," *Journal of Field Robotics*, vol. 24, no. 5, pp. 357–378, 2007.

[115] J. Kelly and G. S. Sukhatme, "Visual-inertial sensor fusion: Localization, mapping and sensor-to-sensor self-calibration," *The International Journal of Robotics Research*, vol. 30, no. 1, pp. 56–79, 2011.

[116] T. Lupton and S. Sukkarieh, "Visual-inertial-aided navigation for high-dynamic motion in built environments without initial conditions," *IEEE Transactions on Robotics*, vol. 28, no. 1, pp. 61–76, 2011.

[117] S. Shen, N. Michael, and V. Kumar, "Tightly-coupled monocular visual-inertial fusion for autonomous flight of rotorcraft mavs," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2015, pp. 5303–5310.

[118] S. Leutenegger, S. Lynen, M. Bosse, R. Siegwart, and P. Furgale, "Keyframe-based visual–inertial odometry using nonlinear optimization," *The International Journal of Robotics Research*, vol. 34, no. 3, pp. 314–334, 2015.

[119] L. Kneip, M. Chli, and R. Siegwart, "Robust real-time visual odometry with a single camera and an imu," in *Proceedings of the British Machine Vision Conference 2011*. British Machine Vision Association, 2011.

[120] S. Weiss and R. Siegwart, "Real-time metric state estimation for modular vision-inertial systems," in *2011 IEEE international conference on robotics and automation*. IEEE, 2011, pp. 4531–4537.

[121] S. Lynen, M. W. Achtelik, S. Weiss, M. Chli, and R. Siegwart, "A robust and modular multi-sensor fusion approach applied to mav navigation," in *2013 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 2013, pp. 3923–3929.

[122] V. Indelman, S. Williams, M. Kaess, and F. Dellaert, "Information fusion in navigation systems via factor graph based incremental smoothing," *Robotics and Autonomous Systems*, vol. 61, no. 8, pp. 721–738, 2013.

[123] N. Sadeghzadeh-Nokhodberiz and J. Poshtan, "Loosely coupled fusion of camera and inertial sensors for distributed error compensation in strapdown inertial navigation system," *Transactions of the Institute of Measurement and Control*, vol. 38, no. 11, pp. 1283–1297, 2016.

[124] C. Forster, L. Carlone, F. Dellaert, and D. Scaramuzza, "Imu preintegration on manifold for efficient visual-inertial maximum-a-posteriori estimation." Georgia Institute of Technology, 2015.

[125] K. Eckenhoff, Y. Yang, P. Geneva, and G. Huang, "Tightly-coupled visual-inertial localization and 3-d rigid-body target tracking," *IEEE Robotics and Automation Letters*, vol. 4, no. 2, pp. 1541–1548, 2019.

[126] M. Li and A. I. Mourikis, "Optimization-based estimator design for vision-aided inertial navigation," in *Robotics: Science and Systems*. Berlin Germany, 2013, pp. 241–248.

[127] M. Li, "Visual-inertial odometry on resource-constrained systems," Ph.D. dissertation, 2014.

[128] Z. Zhang, A. A. Suleiman, L. Carlone, V. Sze, and S. Karaman, "Visual-inertial odometry on chip: An algorithm-and-hardware co-design approach," in *Proceedings of Robotics: Science and Systems*. Massachusetts Institute of Technology, 2017.

[129] P. D. Groves, *Principles of GNSS, Inertial, and Multisensor Integrated Navigation Systems, Second Edition*, December 2007.

[130] Y. A. Felus and R. C. Burtch, "On symmetrical three-dimensional datum conversion," *GPS solutions*, vol. 13, no. 1, p. 65, 2009.

[131] L. Iozan, J. Collin, O. Pekkalin, J. Hautamäki, J. Takala, and C. Rusu, "Measuring the earth's rotation rate using a low-cost mems gyroscope," in *Proceedings of the 2010 Symposium Gyro Technology, Karlsruhe, Germany*, vol. 2122, 2010.

[132] H. F. Martin, "Overcoming the challenges of low-cost inertial navigation," Ph.D. dissertation, UCL (University College London), 2016.

[133] M. Elhoushi, "Advanced motion mode recognition for portable navigation," Ph.D. dissertation, Queen's University, 2015.

[134] P. Aggarwal, *MEMS-based integrated navigation*. Artech House, 2010.

[135] O. Le Traon, D. Janiaud, S. Muller, P. Bouniol, and .-C. F. Office National d'Etudes et de Recherches Aerospatiales (ONERA), "The via vibrating beam accelerometer: concept and performances," pp. 25–29, 1998.

[136] D. Titterton, J. L. Weston, and J. Weston, *Strapdown inertial navigation technology*. IET, 2004, vol. 17.

[137] P. Aggarwal, Z. Syed, X. Niu, and N. El-Sheimy, "A standard testing and calibration procedure for low cost mems inertial sensors and units," *The Journal of Navigation*, vol. 61, no. 2, pp. 323–336, 2008.

[138] V. Kempe, *Inertial MEMS: principles and practice*. Cambridge University Press, 2011.

[139] H. Martin, P. Groves, M. Newman, and R. Faragher, "A new approach to better low-cost mems imu performance using sensor arrays," in *Proceedings of the 26th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2013)*, 2013, pp. 2125–2142.

[140] Q. Wu, R. Wu, F. Han, and R. Zhang, "A three-stage accelerometer self-calibration technique for space-stable inertial navigation systems," *Sensors*, vol. 18, no. 9, p. 2888, 2018.

[141] O. Särkkä, T. Nieminen, S. Suuriniemi, and L. Kettunen, "A multi-position calibration method for consumer-grade accelerometers, gyroscopes, and magnetometers to field conditions," *IEEE Sensors Journal*, vol. 17, no. 11, pp. 3470–3481, 2017.

[142] C. Ren, Q. Liu, and T. Fu, "A novel self-calibration method for mimu," *IEEE Sensors Journal*, vol. 15, no. 10, pp. 5416–5422, 2015.

[143] S. Khankalantary, S. Ranjbaran, and S. Ebadollahi, "Simplification of calibration of low-cost mems accelerometer and its temperature compensation without accurate laboratory equipment," *Measurement Science and Technology*, vol. 32, no. 4, p. 045102, 2021.

[144] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, 2004.

[145] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[146] M. H. Jones, *Calibration Checkerboard Collection*, (Accessed: April 25, 2021). [Online]. Available: https://markhedleyjones.com/projects/calibration-checkerboard-collection

[147] S. Xiao, W. Tao, and H. Zhao, "A flexible fringe projection vision system with extended mathematical model for accurate three-dimensional measurement," *Sensors*, vol. 16, no. 5, p. 612, 2016.

[148] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000.

[149] T. Photometrics, *Rolling vs Global Shutter*, (Accessed: December 15, 2021). [Online]. Available: https://www.photometrics.com/learn/advanced-imaging/rolling-vs-global-shutter

[150] D. Schubert, N. Demmel, L. von Stumberg, V. Usenko, and D. Cremers, "Rolling-shutter modelling for direct visual-inertial odometry," in *IEEE International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2019, pp. 2462–2469.

[151] J. Vautherin, S. Rutishauser, K. Schneider-Zapp, H. F. Choi, V. Chovancova, A. Glass, and C. Strecha, "Photogrammetric accuracy and modeling of rolling shutter cameras." *ISPRS Annals of Photogrammetry, Remote Sensing & Spatial Information Sciences*, vol. 3, no. 3, 2016.

[152] C. Chen, H. Zhu, M. Li, and S. You, "A review of visual-inertial simultaneous localization and mapping from filtering-based and optimization-based perspectives," *Robotics*, vol. 7, no. 3, p. 45, 2018.

[153] G. Cioffi and D. Scaramuzza, "Tightly-coupled fusion of global positional measurements in optimization-based visual-inertial odometry," in *2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2020, pp. 5089–5095.

[154] T. H. Nguyen, T.-M. Nguyen, and L. Xie, "Range-focused fusion of camera-imu-uwb for accurate and drift-reduced localization," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 1678–1685, 2021.

[155] Y. Song and L.-T. Hsu, "Tightly coupled integrated navigation system via factor graph for uav indoor localization," *Aerospace Science and Technology*, vol. 108, p. 106370, 2021.

[156] D. Bousdar Ahmed and E. Munoz Diaz, "Loose coupling of wearable-based inss with automatic heading evaluation," *Sensors*, vol. 17, no. 11, p. 2534, 2017.

[157] S. Sirtkaya, B. Seymen, and A. A. Alatan, "Loosely coupled kalman filtering for fusion of visual odometry and inertial navigation," in *Proceedings of the 16th International Conference on Information Fusion*. IEEE, 2013, pp. 219–226.

[158] A. Angrisano, M. Vultaggio, S. Gaglione, and N. Crocetto, "Pedestrian localization with pdr supplemented by gnss," in *2019 European Navigation Conference (ENC)*. IEEE, 2019, pp. 1–6.

[159] J. Perul and V. Renaudin, "Head: smooth estimation of walking direction with a handheld device embedding inertial, gnss, and magnetometer sensors," *NAVIGATION, Journal of the Institute of Navigation*, vol. 67, no. 4, pp. 713–726, 2020.

[160] D. B. Ahmed, L. E. D. Blanco, and E. M. Diaz, "Performance comparison of wearable-based pedestrian navigation systems in large areas," in *Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2017, pp. 1–7.

[161] T. N. Do and Y. S. Suh, "Gait analysis using floor markers and inertial sensors," *Sensors*, vol. 12, no. 2, pp. 1594–1611, 2012.

[162] M. Placer and S. Kovačič, "Enhancing indoor inertial pedestrian navigation using a shoe-worn marker," *Sensors*, vol. 13, no. 8, pp. 9836–9859, 2013.

[163] J. Hojin, "Pedestrian dead reckoning system using foot kinematic constraint and multiple virtual track technique," Ph.D. dissertation, Seoul National University, 2018.

[164] J. Wahlström, A. Markham, and N. Trigoni, "Footslam meets adaptive thresholding," *IEEE Sensors Journal*, vol. 20, no. 16, pp. 9351–9358, 2020.

[165] V. Mondéjar-Guerra, S. Garrido-Jurado, R. Muñoz-Salinas, M. J. Marín-Jiménez, and R. Medina-Carnicer, "Robust identification of fiducial markers in challenging conditions," *Expert Systems with Applications*, vol. 93, pp. 336–345, 2018.

[166] F. J. Romero-Ramirez, R. Muñoz-Salinas, and R. Medina-Carnicer, "Speeded up detection of squared fiducial markers," *Image and vision Computing*, vol. 76, pp. 38–47, 2018.

[167] S. Garrido-Jurado, R. Munoz-Salinas, F. J. Madrid-Cuevas, and R. Medina-Carnicer, "Generation of fiducial marker dictionaries using mixed integer linear programming," *Pattern Recognition*, vol. 51, pp. 481–491, 2016.

[168] M. J. Floor-Westerdijk, H. M. Schepers, P. H. Veltink, E. H. van Asseldonk, and J. H. Buurke, "Use of inertial sensors for ambulatory assessment of center-of-mass displacements during walking," *IEEE transactions on biomedical engineering*, vol. 59, no. 7, pp. 2080–2084, 2012.

[169] L. Tesio and V. Rota, "The motion of body center of mass during walking: a review oriented to clinical applications," *Frontiers in neurology*, vol. 10, p. 999, 2019.

[170] P. G. Adamczyk and A. D. Kuo, "Redirection of center-of-mass velocity during the step-to-step transition of human walking," *Journal of Experimental Biology*, vol. 212, no. 16, pp. 2668–2678, 2009.

[171] S. A. Gard, S. C. Miff, and A. D. Kuo, "Comparison of kinematic and kinetic methods for computing the vertical motion of the body center of mass during walking," *Human movement science*, vol. 22, no. 6, pp. 597–610, 2004.

[172] J. Rehder, J. Nikolic, T. Schneider, T. Hinzmann, and R. Siegwart, "Extending kalibr: Calibrating the extrinsics of multiple imus and of individual axes," in *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, pp. 4304–4311.

[173] P. Furgale, J. Rehder, and R. Siegwart, "Unified temporal and spatial calibration for multi-sensor systems," in *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, 2013, pp. 1280–1286.

[174] J. S. Furtado, H. H. Liu, G. Lai, H. Lacheray, and J. Desouza-Coelho, "Comparative analysis of optitrack motion capture systems," in *Advances in Motion Sensing and Control for Robotic Applications*. Springer, 2019, pp. 15–31.

[175] P. Merriaux, Y. Dupuis, R. Boutteau, P. Vasseur, and X. Savatier, "A study of vicon system positioning performance," *Sensors*, vol. 17, no. 7, p. 1591, 2017.

[176] OptiTrack, *OptiTrack - Hardware*, (Accessed: Feburary 25, 2021). [Online]. Available: https://optitrack.com/cameras/

[177] R. M. Estep, "Development of a real-time pedestrian localisation system for a warehouse environment," Master's thesis, University of Canterbury, 2016.

[178] C. Villien, A. Frassati, and B. Flament, "Evaluation of an indoor localization engine," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2019, pp. 1–8.

[179] L. Zheng, W. Zhou, W. Tang, X. Zheng, A. Peng, and H. Zheng, "A 3d indoor positioning system based on low-cost mems sensors," *Simulation Modelling Practice and Theory*, vol. 65, pp. 45–56, 2016.

[180] M. N. Muhammad, Z. Salcic, I. Kevin, and K. Wang, "Indoor pedestrian tracking using consumer-grade inertial sensors with pztd heading correction," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5164–5172, 2018.

[181] V. Radu, "Multimodal sensing for robust and energy-efficient context detection with smart mobile devices," Ph.D. dissertation, University of Edinburgh, 2017.

[182] L.-F. Shi, Y.-L. Zhao, G.-X. Liu, S. Chen, Y. Wang, and Y.-F. Shi, "A robust pedestrian dead reckoning system using low-cost magnetic and inertial sensors," *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 8, pp. 2996–3003, 2018.

[183] S. Qiu, Z. Wang, H. Zhao, K. Qin, Z. Li, and H. Hu, "Inertial/magnetic sensors based pedestrian dead reckoning by means of multi-sensor fusion," *Information Fusion*, vol. 39, pp. 108–119, 2018.

[184] J. Rantanen, M. Mäkelä, L. Ruotsalainen, and M. Kirkko-Jaakkola, "Motion context adaptive fusion of inertial and visual pedestrian navigation," in *International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2018, pp. 206–212.

[185] J. Chen, G. Ou, A. Peng, L. Zheng, and J. Shi, "An ins/wifi indoor localization system based on the weighted least squares," *Sensors*, vol. 18, no. 5, p. 1458, 2018.

[186] P. Chen, Y. Kuang, and X. Chen, "A uwb/improved pdr integration algorithm applied to dynamic indoor positioning for pedestrians," *Sensors*, vol. 17, no. 9, p. 2065, 2017.

[187] M. Kwak, Y. Park, J. Kim, J. Han, and T. Kwon, "An energy-efficient and lightweight indoor localization system for internet-of-things (iot) environments,"

Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2, no. 1, pp. 1–28, 2018.

[188] P. Tarrío, M. Cesana, and A. Redondi, "Energy-accuracy trade-offs for hybrid localization using rss and inertial measurements in wireless sensor networks," *Ad hoc networks*, vol. 11, no. 6, pp. 1874–1889, 2013.

[189] H. Aly, A. Basalamah, and M. Youssef, "Accurate and energy-efficient gps-less outdoor localization," *ACM Transactions on Spatial Algorithms and Systems (TSAS)*, vol. 3, no. 2, pp. 1–31, 2017.

[190] K. Zheng, H. Wang, H. Li, W. Xiang, L. Lei, J. Qiao, and X. S. Shen, "Energy-efficient localization and tracking of mobile devices in wireless sensor networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 3, pp. 2714–2726, 2016.

[191] S. Han, T. Li, C. Qian, D. Leith, A. K. Mok, and S. S. Lam, "Hartfi: an energy-efficient localization system," in *Proceedings of the 2nd ACM SIGCOMM workshop on Green networking*, 2011, pp. 25–30.

[192] J. A. Williamson, "Low-power system design for human-borne sensing," Ph.D. dissertation, University of Colorado, 2016.

[193] A. Guidara, F. Derbel, G. Fersi, S. Bdiri, and M. B. Jemaa, "Energy-efficient on-demand indoor localization platform based on wireless sensor networks using low power wake up receiver," *Ad Hoc Networks*, vol. 93, p. 101902, 2019.

[194] J. Xie, D. Wen, L. Liang, Y. Jia, L. Gao, and J. Lei, "Evaluating the validity of current mainstream wearable devices in fitness tracking under various physical activities: comparative study," *JMIR mHealth and uHealth*, vol. 6, no. 4, p. e94, 2018.

[195] J. Pino-Ortega, C. D. Gómez-Carmona, and M. Rico-González, "Accuracy of xiaomi mi band 2.0, 3.0 and 4.0 to measure step count and distance for physical activity and healthcare in adults over 65 years," *Gait & Posture*, vol. 87, pp. 6–10, 2021.

[196] Z. Deng, P. Wang, D. Yan, and K. Shang, "Foot-mounted pedestrian navigation method based on gait classification for three-dimensional positioning," *IEEE Sensors Journal*, vol. 20, no. 4, pp. 2045–2055, 2019.

[197] M. Ji, J. Liu, X. Xu, and Z. Lu, "The improved 3d pedestrian positioning system based on foot-mounted inertial sensor," *IEEE Sensors Journal*, 2020.

[198] J. Wahlström, I. Skog, F. Gustafsson, A. Markham, and N. Trigoni, "Zero-velocity detection—a bayesian approach to adaptive thresholding," *IEEE Sensors Letters*, vol. 3, no. 6, pp. 1–4, 2019.

[199] B. Wagstaff and J. Kelly, "Lstm-based zero-velocity detection for robust inertial navigation," in *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, 2018, pp. 1–8.

[200] M. Mäkelä, J. Rantanen, M. Kirkko-Jaakkola, and L. Ruotsalainen, "Context recognition in infrastructure-free pedestrian navigation—toward adaptive filtering algorithm," *IEEE Sensors Journal*, vol. 18, no. 17, pp. 7253–7264, 2018.

[201] J. Rantakokko, P. Strömbäck, and P. Anderson, "Foot-and knee-mounted ins for firefighter localization," in *Proceedings of the 2014 International Technical Meeting of the Institute of Navigation*, 2014, pp. 145–153.

[202] R. Feliz Alonso, E. Zalama Casanova, and J. Gómez García-Bermejo, "Pedestrian tracking using inertial sensors," 2009.

[203] B. Wagstaff, V. Peretroukhin, and J. Kelly, "Robust data-driven zero-velocity detection for foot-mounted inertial navigation," *IEEE Sensors Journal*, vol. 20, no. 2, pp. 957–967, 2019.

[204] M. Liberty, *Joulescope*, (Accessed: May 25, 2021). [Online]. Available: https://www.joulescope.com/

[205] Qoitech, *Otii Arc*, (Accessed: May 25, 2021). [Online]. Available: https://www.qoitech.com/otii/

[206] LowPowerLab, *CurrentRanger*, (Accessed: May 25, 2021). [Online]. Available: https://lowpowerlab.com/

[207] Z. Chen, X. Pan, C. Chen, and M. Wu, "Contrastive learning of zero-velocity detection for pedestrian inertial navigation," *IEEE Sensors Journal*, 2021.

[208] M. Elhoushi, J. Georgy, A. Noureldin, and M. J. Korenberg, "A survey on approaches of motion mode recognition using sensors," *IEEE Transactions on intelligent transportation systems*, vol. 18, no. 7, pp. 1662–1686, 2016.

[209] P. Kasebzadeh, K. Radnosrati, G. Hendeby, and F. Gustafsson, "Joint pedestrian motion state and device pose classification," *IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 8, pp. 5862–5874, 2019.

[210] M. Cornacchia, K. Ozcan, Y. Zheng, and S. Velipasalar, "A survey on activity detection and classification using wearable sensors," *IEEE Sensors Journal*, vol. 17, no. 2, pp. 386–403, 2016.

[211] A. M. Khan, Y.-K. Lee, S. Y. Lee, and T.-S. Kim, "A triaxial accelerometer-based physical-activity recognition via augmented-signal features and a hierarchical recognizer," *IEEE transactions on information technology in biomedicine*, vol. 14, no. 5, pp. 1166–1172, 2010.

[212] J. Brusey, R. Rednic, E. I. Gaura, J. Kemp, and N. Poole, "Postural activity monitoring for increasing safety in bomb disposal missions," *Measurement Science and Technology*, vol. 20, no. 7, p. 075204, 2009.

[213] N. Ravi, N. Dandekar, P. Mysore, and M. L. Littman, "Activity recognition from accelerometer data," in *Aaai*, vol. 5, no. 2005. Pittsburgh, PA, 2005, pp. 1541–1546.

[214] Y.-P. Chen, J.-Y. Yang, S.-N. Liou, G.-Y. Lee, and J.-S. Wang, "Online classifier construction algorithm for human activity detection using a tri-axial accelerometer," *Applied Mathematics and Computation*, vol. 205, no. 2, pp. 849–860, 2008.

[215] J. R. Kwapisz, G. M. Weiss, and S. A. Moore, "Activity recognition using cell phone accelerometers," *ACM SigKDD Explorations Newsletter*, vol. 12, no. 2, pp. 74–82, 2011.

[216] P. D. Groves, H. Martin, K. Voutsis, D. Walter, and L. Wang, "Context detection, categorization and connectivity for advanced adaptive integrated navigation," in *Proceedings of the 26th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS+ 2013)*, 2013, pp. 1039–1056.

[217] L. Bao and S. S. Intille, "Activity recognition from user-annotated acceleration data," in *International conference on pervasive computing*. Springer, 2004, pp. 1–17.

[218] J. Yang, "Toward physical activity diary: motion recognition using simple acceleration features with mobile phones," in *Proceedings of the 1st international workshop on Interactive multimedia for consumer electronics*, 2009, pp. 1–10.

[219] T. M. Nijsen, R. M. Aarts, P. J. Cluitmans, and P. A. Griep, "Time-frequency analysis of accelerometry data for detection of myoclonic seizures," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 5, pp. 1197–1203, 2010.

[220] M. Susi, D. Borio, and G. Lachapelle, "Accelerometer signal features and classification algorithms for positioning applications," in *Proceedings of the 2011 International Technical Meeting of The Institute of Navigation*, 2011, pp. 158–169.

[221] K. Frank, M. J. Vera-Nadales, P. Robertson, and M. Angermann, "Reliable real-time recognition of motion related human activities using mems inertial sensors," in *Proceedings of the 23rd International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS 2010)*, 2010, pp. 2919–2932.

[222] T. Choudhury, G. Borriello, S. Consolvo, D. Haehnel, B. Harrison, B. Hemingway, J. Hightower, P. Pedja, K. Koscher, A. LaMarca *et al.*, "The mobile sensing platform: An embedded activity recognition system," *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 32–41, 2008.

[223] K. Frank, E. M. Diaz, P. Robertson, and F. Sanchez, "Bayesian recognition of safety relevant motion activities with inertial sensors and barometer," in *Proceedings of IEEE/ION PLANS 2014*, 2014, pp. 174–184.

[224] S. Theodoridis, "Konstantinos koutroumbas pattern recognition," in *Beijing: Mechanical*. Industrial Press, 2003, pp. 163–205.

[225] Z. Sun, X. Mao, W. Tian, and X. Zhang, "Activity classification and dead reckoning for pedestrian navigation with wearable sensors," *Measurement science and technology*, vol. 20, no. 1, p. 015203, 2008.

[226] O. Z. Maimon and L. Rokach, *Data mining with decision trees: theory and applications*. World scientific, 2014, vol. 81.

[227] C. Chien and G. J. Pottie, "A universal hybrid decision tree classifier design for human activity classification," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 1065–1068.

[228] J. Yin, Q. Yang, and J. J. Pan, "Sensor-based abnormal human-activity detection," *IEEE Transactions on Knowledge and Data Engineering*, vol. 20, no. 8, pp. 1082–1090, 2008.

[229] M. J. Korenberg, "A robust orthogonal algorithm for system identification and time-series analysis," *Biological cybernetics*, vol. 60, no. 4, pp. 267–276, 1989.

[230] M. J. Korenberg, R. David, I. W. Hunter, and J. E. Solomon, "Automatic classification of protein sequences into structure/function groups via parallel cascade identification: a feasibility study," *Annals of biomedical engineering*, vol. 28, no. 7, pp. 803–811, 2000.

[231] Y.-L. Hsu, S.-C. Yang, H.-C. Chang, and H.-C. Lai, "Human daily and sport activity recognition using a wearable inertial sensor network," *IEEE Access*, vol. 6, pp. 31 715–31 728, 2018.

[232] L. Gao, A. Bourke, and J. Nelson, "Evaluation of accelerometer based multi-sensor versus single-sensor activity recognition systems," *Medical engineering & physics*, vol. 36, no. 6, pp. 779–785, 2014.

[233] M. Li, J. Lu, Z. Chen, and K. Amine, "30 years of lithium-ion batteries," *Advanced Materials*, vol. 30, no. 33, p. 1800561, 2018.

[234] "Tello," *RYZE*. Accessed Dec 21, 2021 [Online]. [Online]. Available: https://www.ryzerobotics.com/tello

[235] L. E. Díez, A. Bahillo, J. Otegui, and T. Otim, "Step length estimation methods based on inertial sensors: A review," *IEEE Sensors Journal*, vol. 18, no. 17, pp. 6908–6926, 2018.

[236] J.-O. Nilsson, D. Zachariah, I. Skog, and P. Händel, "Cooperative localization by dual foot-mounted inertial sensors and inter-agent ranging," *EURASIP Journal on Advances in Signal Processing*, vol. 2013, no. 1, pp. 1–17, 2013.

[237] P. Zhang, R. Chen, Y. Li, X. Niu, L. Wang, M. Li, and Y. Pan, "A localization database establishment method based on crowdsourcing inertial sensor data and

quality assessment criteria," *IEEE Internet of Things Journal*, vol. 5, no. 6, pp. 4764–4777, 2018.

[238] F. Zampella, A. R. J. Ruiz, and F. S. Granja, "Indoor positioning using efficient map matching, rss measurements, and an improved motion model," *IEEE Transactions on Vehicular Technology*, vol. 64, no. 4, pp. 1304–1317, 2015.

[239] D. Prutchi, *Exploring Ultraviolet Photography: Bee Vision, Forensic Imaging, and Other NearUltraviolet Adventures with Your DSLR*. Amherst Media, 2016.

[240] E. Urtans and A. Nikitenko, "Active infrared markers for augmented and virtual reality," *markers*, vol. 9, p. 10, 2016.

[241] Y. Nakazato, M. Kanbara, and N. Yokoya, "Localization system for large indoor environments using invisible markers," in *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, 2008, pp. 295–296.

[242] Y. Kamiya, Y. Gu, and S. Kamijo, "Indoor positioning in large shopping mall with context based map matching," in *2019 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2019, pp. 1–6.

[243] Y. Chen, F. Wu, W. Shuai, N. Wang, R. Chen, and X. Chen, "Kejia robot–an attractive shopping mall guider," in *International Conference on Social Robotics*. Springer, 2015, pp. 145–154.