Ollscoil na hÉireann, Corcaigh

National University of Ireland, Cork

Data Fusion for Human Motion Tracking with Multimodal Sensing

Volume 1 of 1

Thesis presented by

Mariusz P. Wilk, BEng (Hons)

orcid.org/0000-0002-2780-1110

for the degree of

Doctor of Philosophy

University College Cork

School of Electrical and Electronic Engineering

Head of School Dr Jorge Oliveira

Supervisors: Dr Brendan O'Flynn, Dr Michael Walsh

2020

This thesis is dedicated to my beautiful wife Jocelyn and our son Patrick.

Table of Contents

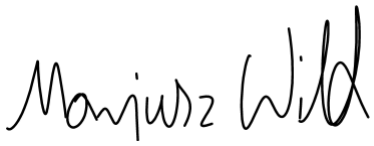
List of Figures

List of Tables

Declaration

This is to certify that the work I am submitting is my own and has not been submitted for another degree, either at University College Cork or elsewhere. All external references and sources are clearly acknowledged and identified within the contents. I have read and understood the regulations of University College Cork concerning plagiarism.

Where other sources of information have been used, they have been acknowledged.

Signature: 

Date: 7.05.2020

Acknowledgements

I would like to thank all the people who helped me complete the work described in this thesis.

First and foremost, I am very thankful to Science Foundation Ireland and its SFI Centre CONNECT (13/RC/2077) for their support. This project would have never existed without it. Secondly, I would like to thank members of my thesis committee who have always supported me and never declined my requests for help or guidance. I am particularly grateful to my supervisors. Dr Alan Mathewson offered an unparalleled mentorship that helped me see my own future. I will always be grateful for that. Secondly, I have to thank Dr Brendan O'Flynn for his never-ending patience to me and my ideas. He has always considered my requests and often let me explore various areas of science, some of which were not strictly related to this thesis. His support has helped me become a better researcher. I am also very thankful to Dr Michael Walsh who has offered a significant amount of guidance and mentorship on many technical aspects of my research work. I would also like to extend many thanks to Dr Andrew O'Connell and Ray Ó Cinnéide from CONNECT who, along with Dr Brendan O'Flynn, helped me become a much better communicator of science by facilitating my participation in numerous outreach activities. I owe special thanks for Professor Cian Ó Mathúna for his support of my activities. I am grateful to the members of the Wireless Sensor Networks who always support each other. I am also grateful to all the support staff in various departments in Tyndall National Institute whose work was vital in helping me complete this thesis. I owe particular thanks to members of the IT Department who always responded to my requests for technical assistance. Also, I would like to express gratitude to my beautiful wife Jocelyn and our son Patrick for their mostly unending patience and support. Last but not least, I am thankful to all people who supported me along the way whose names I have not mentioned.

List of Abbreviations

AA	Alphabetised Abbreviations
BLE	Bluetooth Low Energy
CMOS	Complementary Metal-Oxide Semiconductor
DTM	Demographic Transition Model
FoV	Field-of-View
GA	Gaussian Approximation
HCI	Human Computer Interface
HMD	Head Mounted Display
HW	Hardware
IC	Integrated Circuit
ICT	Information Communications Technology
IMU	Inertial Measurement Unit
IoT	Internet of Things
IR	InfraRed
LED	Light Emitting Diode
LI	Linear Interpolation
LM	Levenberg-Marquardt
LOS	Line-Of-Sight
MCU	Micro-Controller Unit
MEMS	Micro-Electro-Mechanical Systems
MP	Mega Pixel
PnP	Perspective-n-Point
PSF	Point Spread Function
PWM	Pulse Width Modulation
RF	Radio Frequency
RMSE	Root Mean Square Error
SLAM	Simultaneous Localisation And Mapping
SLI	Simplified Linear Interpolation
SOA	State-Of-the-Art
ST	Strength Training
SW	Software
TOF	Tome-Of-Flight
UWB	Ultra-Wide Band
VR	Virtual Reality
WP	Wearable Platform
1-D	1-Dimensional
2-D	2-Dimensional
3-D	3-Dimensional
6-DOF	6-Degree-Of-Freedom

Abstract

Multimodal sensor fusion is a common approach in the design of many motion tracking systems. It is based on using more than one sensor modality to measure different aspects of a phenomenon and capture more information about it than what would be available otherwise from a single sensor. Multimodal sensor fusion algorithms often leverage the complementary nature of the different modalities to compensate for shortcomings of the individual sensor modalities. This approach is particularly suitable for low-cost and highly miniaturised wearable human motion tracking systems that are expected to perform their function with limited resources at their disposal (energy, processing power, etc.). Opto-inertial motion trackers are some of the most commonly used approaches in this context. These trackers fuse the sensor data from vision and Inertial Motion Unit (IMU) sensors to determine the 3-Dimensional (3-D) pose of the given body part, i.e. its position and orientation. The continuous advances in the State-Of-the-Art (SOA) in camera miniaturisation and efficient point detection algorithms along with the more robust IMUs and increasing processing power in a shrinking form factor, make it increasingly feasible to develop a low-cost, low-power, and highly miniaturised wearable smart sensor human motion tracking system. It incorporates these two sensor modalities. In this thesis, a multimodal human motion tracking system is presented that builds on these developments. The proposed system consists of a wearable smart sensor system, referred to as Wearable Platform (WP), which incorporates the two sensor modalities, i.e. monocular camera (optical) and IMU (motion). The WP operates in conjunction with two optical points of reference embedded in the ambient environment to enable positional tracking in that environment. In addition, a novel multimodal sensor fusion algorithm is proposed which uses the complementary nature of the vision and IMU sensors in conjunction with the two points of reference in the ambient environment, to determine the 3-D pose of the WP in a novel and computationally efficient way.

To this end, the WP uses a low-resolution camera to track two points of reference; specifically two Infrared (IR) LEDs embedded in the wall. The geometry that is formed between the WP and the IR LEDs, when complemented by the angular rotation measured by the IMU, simplifies the mathematical formulations involved in the computing the 3-D pose, making them compatible with the resource-constrained microprocessors used in such wearable systems. Furthermore, the WP is coupled with the two IR LEDs via a radio link to control their intensity in real-time. This enables the novel subpixel point detection algorithm to maintain its highest accuracy, thus

increasing the overall precision of the pose detection algorithm. The resulting 3-D pose can be used as an input to a higher-level system for further use.

One of the potential uses for the proposed system is in sports applications. For instance, it could be particularly useful for tracking the correctness of executing certain exercises in Strength Training (ST) routines, such as the barbell squat. Thus, it can be used to assist professional ST coaches in remotely tracking the progress of their clients, and most importantly ensure a minimum risk of injury through real-time feedback. Despite its numerous benefits, the modern lifestyle has a negative impact on our health due to an increasingly sedentary lifestyle that it involves. The human body has evolved to be physically active. Thus, these lifestyle changes need to be offset by the addition of regular physical activity to everyday life, of which ST is an important element.

This work describes the following novel contributions:

- A new multimodal sensor fusion algorithm for 3-D pose detection with reduced mathematical complexity for resource-constrained platforms
- A novel system architecture for efficient 3-D pose detection for human motion tracking applications
- A new subpixel point detection algorithm for efficient and precise point detection at reduced camera resolution
- A new reference point estimation algorithm for finding locations of reference points used in validating subpixel point detection algorithms
- A novel proof-of-concept demonstrator prototype that implements the proposed system architecture and multimodal sensor fusion algorithm

1 Introduction

1.1 Research Motivation and Rationale

Multimodal sensor fusion is a concept borrowed from nature. Most living organisms use it to survive; including humans. The key property of multimodal sensor fusion systems is data complementarity wherein various sensory modalities are used to sense different aspects of the same phenomenon in order to learn more about it than what would be possible with a single modality [1]. An example of a multimodal sensor fusion system in nature includes the human's audio-visual system. Humans use these two complementary sensor modalities to better understand the environment around them. For example, when one attends a lecture at a university, most information is conveyed through the speaker's voice, but one's vision is used to complement it by providing clues on speaker's body language, lecture notes, etc. Multimodal sensor fusion is used in science and engineering in the same way. In this field, sensors of various modalities are used to gather more information about the given phenomenon than what would be possible otherwise using a single data source. It can be used in various application spaces, one of which is human motion tracking, the focus of this work.

With continued developments in increasing the computing power and shrinking the size of electronic devices, it becomes increasingly feasible to use technology for human motion tracking leveraging affordable, low-power, and highly miniaturised wearable devices. These wearable devices can be based on smart sensor systems that incorporate a number of sensor modalities along with processing and telecommunications capabilities. These can be programmed with intelligent algorithms based on multimodal sensor fusion to efficiently perform the human motion tracking function, i.e. by using various sensor types to complement the weaknesses of individual sensors through data combination. Such tracking systems can be used to track human motion and give feedback in real time to the user or their coach (in, say, a sports application). This thesis describes the development and validation of such a motion tracking system. The review of the current State-Of-The-Art (SOA) showed that the knowledge in low-power motion tracking systems advances at a significant pace; with the evidence of multiple streams of research being reported in the literature, as described in Chapter 2. Moreover, the SOA review helped to identify a gap in knowledge, which, in turn, was used to formulate the hypothesis of this work, described in Section 2.3. The work described in this thesis was aimed at proving this hypothesis.

To this end, the proposed system incorporates a wearable device that houses a monocular camera sensor and an Inertial Motion Unit (IMU) sensor. It runs the novel multimodal sensor fusion algorithm, developed and described in this thesis, which determines the position and orientation of the wearable device in space using the information from the IMU and camera that keeps track of two points of reference in the ambient environment to enable position tracking. Due to its small size, the wearable device can be attached to various body parts to track their motion over time. The proposed system may be used in many application spaces that require the ability to perform motion tracking.

One such application is in sports and fitness monitoring. Such a system may be used to track the motion of individuals engaged in performing certain physical exercise routines. Thus, it can be used safety and performance monitoring in real-time. Safety and performance tracking in Strength Training (ST) is an example of a specific application wherein the proposed system may be particularly beneficial. It was used as a demonstrator for the novel motion tracking technologies developed as the ST is recommended as an addition to regular physical activity [2]. Such technology can be used to track the motion of the human body and ensure that the given exercise is executed correctly, thus minimising the risk of injury through the supervision of a "virtual coach" [3]. The technology described in this thesis is focused on enabling accurate positioning of the human body while exercising as part of the Strength Training (ST) regime.

1.1.1 *Current State-of-the-Art - Summary*

Human motion tracking is a popular topic in the research community. The interest in monitoring human motion dates back as far as human history does. Its modern use began in the 20th century [4]. It was sparked by the developments in photography and later the invention and widespread use of electronic systems, such as semiconductors and computers in general. These breakthroughs enabled researchers to use sensors and computers to capture and track the motion of the human body over time. Initially, such systems were highly limited in functionality, which often bulky, and could be used only in highly controlled laboratory conditions for specific purposes, such as head positioning system for Head Mounted Devices (HMD) proposed by Sutherland et al. [5] . The continuing advances in the SOA in system miniaturisation and increasing computing power were some of the enabling factors for new application spaces involving human motion tracking technology [6]. An example of one of the first real-world applications of human motion tracking technology includes the movie making industry, where

the motion of human actors was captured and used to control animated characters. These motion trackers are usually based on using multiple cameras with markers attached to the human body. Such systems are usually versatile and offer high accuracy [7]. However, they are expensive and require a relatively complicated setup and significant computational power even for offline data processing. There exist low-cost alternatives, but their performance does not match that of the expensive high-end systems in terms of accuracy [8, 9]. In recent years, wearable wireless smart sensor systems have been increasingly used for human motion tracking applications, both in unimodal and multimodal sensor configurations [10, 11]. Whereas unimodal approaches rely on a single sensor modality, the multimodal systems use more than one sensor modality, e.g. opto-inertial trackers use IMUs and Vision sensor modalities. The IMU based systems tend to be the most widely used unimodal approaches to motion tracking (despite arguably being multimodal devices, because they typically consist of an accelerometer, gyroscope and magnetometer, i.e. various sensor modalities). They have numerous advantages, such as: low cost, small-form-factor, energy efficiency or accurate orientation tracking. However, despite their advantages, IMUs suffer from problems that prevent them from reliably tracking the absolute position over extended periods of time, such as drifts or susceptibility to disturbances in magnetic field. An IMU can track the position accurately only for short periods of time; before its position estimate drifts unacceptably far away from the true value. In order to counteract this limitation, the IMUs are often used with other sensors in multimodal setups. An increasingly popular multimodal sensor configuration found in the literature includes opto-inertial trackers, i.e. systems that integrate IMUs with vision sensors. These two sensor modalities complement the weaknesses of each of the individual sensor modalities. Whereas vision sensors tend to perform poorly in the presence of occlusions and uncontrolled lighting conditions, the IMUs are robust under such conditions. On the other hand, the IMU cannot be used to reliably and robustly measure absolute positions while the vision sensors can. Likewise, orientation measurement using a vision sensor is often difficult to do and has high processing power requirements while this is easily achieved using the IMUs. Moreover, a monocular low-cost vision sensor is not particularly well suited to performing 3-D pose detection, i.e. determining camera's position and orientation in the environment. While it can be achieved using such a single low-cost camera, it is generally a difficult and often impractical task in the context of low-cost and small form-factor, wearable, systems. Therefore, a combination of these two sensor modalities is beneficial, because such opto-inertial trackers can be used to determine the complete 3-Dimensional (3-D) pose of the

object being tracked, i.e. the position and orientation in three dimensions, using low cost components [11-13].

1.1.2 *Gap in the State-of-the-Art - Summary*

With the continuing advances of the SOA in the miniaturisation and integration of electronic devices, along with embedded algorithms, it becomes increasingly feasible to perform human motion tracking using low-power and small-form-factor wireless wearable systems. The miniaturisation of vision sensors, and increase in computing power, as well as decreasing power consumption of electronic components, are some of the key enabling factors. For instance, the lens-less vision sensors significantly reduce the size of the regular camera by effectively removing the lens, usually the largest component in a vision sensor [14, 15]. Although these sensors cannot capture images with the same level of detail as the traditional cameras with lenses, a sufficient amount of information can be extracted to perform point tracking to enable accurate positioning [16].

Li et al. and Maereg et al. [12, 13] showed that the 3-D pose of an object can be determined by combining a monocular camera with an IMU in a computationally simplified way, as compared to the Simultaneous Localisation and Mapping (SLAM) and Perspective-n-Point (PnP) methods that are normally used in monocular pose estimation and tracking systems [17, 18]. Both methods show that the combination of an IMU with a camera that tracks two points of reference can be used to efficiently compute the 3-D pose. However, both systems are "outside-in" systems, i.e. the cameras are not embedded in the moving device, which adversely affects their cost and thus the scalability of such a system. On the other hand, the IS-1500 is the most accurate "inside-out" opto-inertial tracker in SOA; with a typical accuracy in positional tracking of 2 mm [11]. Although this system achieves the best performance in a small form factor with both modalities embedded in the wearable device, it has higher computational requirements. It also requires at least four points of reference in the environment.

The developments in camera miniaturisation and opto-inertial motion tracking systems show the potential for lowering the cost and size of wearable inside-out, opto-inertial, human motion tracking systems, such as the IS-1500. The affordability of such systems can be increased by simplifying the overall system complexity and decreasing its computational requirements. To this end, a low-cost camera can be used to track two known points of reference in the ambient environment, similarly to the methods proposed by Li et al. and Maereg et al [12, 13]. However,

the monocular camera can be incorporated in the wearable device along with the IMU in a similar way to that of the IS-1500. Therefore, the wearable tracker can be used to track two known points of reference in the environment. The information obtained by the camera can be complemented by the data from the IMU to determine the 3-D pose of the wearable device. The information from the two complementing sensor modalities can be fused together so as to significantly simplify the mathematical calculations involved in the pose detection algorithm.

These advances are some of the contributing factors to allow addressing a gap in the SOA that exists in this research area. The scientific literature does not show the evidence of an extensive research work aimed at exploring this area. The literature suggests that the research community tends to be focused on exploring different methodologies. Thus, this gap exposes an underexplored research area and suggests the potential directions for further research activities.

1.2 Contribution and Organisation of this Thesis

The aim of the work presented in this thesis is to develop, validate, and demonstrate a proof-of-concept prototype of a wearable human motion tracking system for various application spaces, including sports applications focused on strength and conditioning training. The key objectives of the proposed system were to be able to determine the 3-D pose using a highly miniaturised, resource-constrained, wearable device in the context of low cost, limited processing power and energy consumption and to provide real-time feedback on body motion to the user.

The proposed system consists of two main aspects, the data acquisition system itself and the embedded algorithms required to determine positioning based on sensor data fusion. Firstly, the proposed novel data acquisition system architecture ensures that the system can perform motion tracking in the context of low cost and low-power wearable systems. To this end, the novel wearable motion tracking device incorporates two sensor modalities that complement each other's weaknesses, i.e. a monocular low-cost vision sensor and an IMU. The camera, embedded in the wearable device, is used to track two known points of reference in the ambient environment, i.e. Infrared (IR) Light Emitting Diodes (LED).

Secondly, this work describes a novel sensor fusion algorithm that uses two sensor modalities to directly compute the 3-D pose, i.e. position and orientation, of the wearable device in space. The complementary nature of the vision and inertial sensor modalities along with the proposed system architecture are leveraged to minimise the computational complexity of the 3-D pose estimation

calculations. The reduced computational complexity of the algorithms involved is the main enabling factor for human motion tracking applications using affordable, highly miniaturised, wearable wireless smart sensor systems.

The proposed algorithms reduce the computational requirements in several ways. From a mathematical point of view, the image processing algorithms involved in extracting information form the camera images are amongst the greatest challenges in the context of low power miniaturised wearable devices. These requirements have traditionally made the consideration of using wearable computer vision in this context generally prohibitive. However, our proposed approach tackles these problems since the complexity of the image processing algorithms depends on what information is to be extracted from the image frames, the proposed system architecture significantly simplifies this task; thus, reducing processing requirements. It reduces this task to extracting only two known points from the images, i.e. from an IR LED. Moreover, the camera uses a matching optical IR filter, which further simplifies this process, by suppressing the noise levels. This alone, however, does not completely solve the problem. Point detection algorithms in image processing must process all pixels in every image many times per second, despite the simplicity of the image processing tasks involved in finding the points. For this reason, we propose a novel computationally efficient subpixel point detection algorithm that allows for lowering the camera's resolution while maintaining the precision of the point detection algorithm, without imposing significant overheads.

The coordinates of the two points of reference found in the images are the inputs to the proposed novel sensor fusion algorithm. The two points in the ambient environment along with the camera's principal point and image plane form a set of geometries that our algorithm depends on. The properties of the geometries, such as the similar triangles, are exploited in the mathematical calculations. The missing pieces of information are obtained from the IMU sensor. The IMU provides the rotation angles that fill the critically important gaps in the mathematical model of the 3-D pose calculation.

1.2.1 *Chapter 2 (State-of-the-art)*

This chapter describes the main findings of the review of the relevant scientific literature in positioning and motion tracking. The review is concluded with describing the identified gap in the current SOA and the hypothesis that this thesis addresses.

Under Review: **M. P. Wilk,** M. Walsh, and B. O'Flynn, “Human motion tracking technology for healthier and longer living”, Frontiers, 2020.

1.2.2 *Chapter 3 (Wearable Vision; Point Detection and Tracking)*

In this chapter, the novel subpixel point detection algorithm used for detecting the locations of point centres in images at subpixel level is described. A detailed description of the system modelling and experimental validation is presented. It shows that this algorithm has lower execution time and is more accurate than the relevant alternative methodologies in the literature, for the proposed system architecture. Therefore, its use enables the reduction of the camera resolution without significantly increasing the computational requirements of the wearable device. A novel methodology for reference point estimation for validating and benchmarking the subpixel point algorithms is also described in this chapter.

Under review: **M. P. Wilk**, M. Walsh, and B. O’Flynn, "Extended Efficient Sub-Pixel Point Detection Algorithm for Point Tracking with Low-Power Wearable Camera Systems," IEEE Transactions on Image Processing, 2020.

M. P. Wilk and B. O'Flynn, "Reference Point Estimation Technique for Direct Validation of Subpixel Point Detection Algorithms for Internet of Things," in 2019 30th Irish Signals and Systems Conference (ISSC), 17-18 June 2019, pp. 1-5, DOI:10.1109/ISSC.2019.8904921

1.2.3 *Chapter 4 (Multimodal Sensor Fusion; Monocular 3D Pose Estimation)*

This chapter describes the proposed system architecture and the proposed multimodal sensor fusion algorithm for 3-D pose detection. The detailed mathematical modelling of the algorithm is included. It shows how the mathematical formulations leverage the proposed system architecture and the complementary nature of the two sensor modalities to simplify the calculations. Also, a detailed description of the validation of experimental accuracy validation in laboratory conditions is described.

Under review: **M. P. Wilk**, M. Walsh, and B. O’Flynn, "Multimodal Sensor Fusion for Low-Power Miniaturised Wearable Human Motion Tracking Systems in Sports Applications," IEEE Sensors, 2020.

1.2.4 *Chapter 5 (Embedded Prototype Multimodal Tracking System)*

Chapter 5 gives a description of the development of a proof-of-concept prototype demonstrator system. The system was used to demonstrate the operation of the proposed system architecture and algorithms. An initial experimental evaluation of the prototype was completed. The results showed that the embedded version of the system performed as expected and were consistent with the results of simulations and experimental validation of the pre-prototype, non-wearable, version described in section 4.4 in Chapter 4.

Accepted: **M. P. Wilk,** M. Walsh, and B. O'Flynn, "Embedded Multimodal Opto-Inertial Motion Tracking System", 31th Irish Signals and Systems Conference (ISSC), 2020

1.2.5 *Chapter 6 (Thesis Summary and Conclusions)*

A summary of the thesis is described in this chapter. The main findings of this work, as well as their importance, are included in the form of a brief summary of each chapter. The key contributions of this work are also listed as short bullet points. Finally, suggestions for potential directions of future works are provided which include further development and miniaturisation of the prototype system as well as performance testing with human subjects and exploring potential commercialisation routes for this work.

1.3 Novel Contributions

This work presents the following novel contributions:

- Multimodal sensor fusion algorithm for 3-D pose detection with reduced mathematical complexity
- System architecture for efficient 3-D pose detection for human motion tracking applications
- Subpixel point detection algorithm for efficient and precise point detection at reduced camera resolution
- Reference point estimation algorithm for finding locations of reference points used in validating subpixel point detection algorithms
- A proof-of-concept demonstrator prototype that implements the proposed system architecture and multimodal sensor fusion algorithm

1.4 Publications

Journal Articles; Peer-reviewed, under-review, and in preparation:

- L. Abraham, A. Urru, N. Normani, **M. Wilk**, M. Walsh, and B. O'Flynn, "Hand tracking and gesture recognition using lensless smart sensors," **MDPI Sensors**, vol. 18, no. 9, p. 2834, 2018. (Role: experimental validation of 3D ranging algorithm, offline data analysis, manuscript revision)
- **M. P. Wilk**, M. Walsh, and B. O'Flynn, "Multimodal Sensor Fusion for Low-Power Miniaturised Wearable Human Motion Tracking Systems in Sports Applications," **IEEE Sensors**, 2019, (under review)
- **M. P. Wilk**, M. Walsh, and B. O'Flynn, "Extended Efficient Sub-Pixel Point Detection Algorithm for Point Tracking with Low-Power Wearable Camera Systems," **IEEE Transactions on Image Processing**, 2019, (under review)
- **M. P. Wilk,** M. Walsh, and B. O'Flynn, "Human motion tracking technology for healthier and longer living", **Frontiers**, 2019, (under review)

Conference papers; Peer-reviewed:

- **M. P. Wilk,** M. Walsh, and B. O'Flynn, "Embedded Multimodal Opto-Inertial Motion Tracking System for Sports Applications", **31st Irish Signals and Systems Conference (ISSC)**, 2020, (accepted)
- **M. P. Wilk** and B. O'Flynn, "Reference Point Estimation Technique for Direct Validation of Subpixel Point Detection Algorithms for Internet of Things," in 2019 **30th Irish Signals and Systems Conference (ISSC)**, 17-18 June 2019, pp. 1-5, doi:10.1109/ISSC.2019.8904921
- **M. P. Wilk** and B. O'Flynn, "Miniaturized Low-Power Wearable System for Human Motion Tacking Incorporating Monocular Camera and Inertial Sensor Data Fusion for Health Applications," presented at the **Smart Systems Integration Conference 2019**, Barcelona, Spain, 2019.
- **M. P. Wilk**, J. Torres-Sanchez, S. Tedesco, and B. O. Flynn, "Wearable Human Computer Interface for Control Within Immersive VAMR Gaming Environments Using Data Glove and Hand Gestures," in 2018 **IEEE Games, Entertainment, Media Conference (GEM)**, 15-17 Aug. 2018, pp. 1-9, doi: 10.1109/GEM.2018.8516521

- Lizy Abraham, Andrea Urru, **M. P. Wilk**, Michael Walsh, Brendan O'Flynn, '3D Ranging and Tracking using an Improved Lensless Smart Sensor', **14th *International Conference on Information Processing, IEEE*** *(accepted)*, Dec. 2018, Bangalore, India. (Role: experimental work, manuscript revision)
- **M. P. Wilk**, A. Urru, S. Tedesco, and B. O. Flynn, "Sub-pixel point detection algorithm for point tracking with low-power wearable camera systems: A simplified linear interpolation," in 2017 **28th Irish Signals and Systems Conference (ISSC)**, 20-21 June 2017, pp. 1-6, doi: 10.1109/ISSC.2017.7983629.
- L. Abraham, A. Urru, **M. P. Wilk**, S. Tedesco, M. Walsh, and B. O. Flynn, "Point tracking with lensless smart sensors," in 2017 **IEEE SENSORS**, Oct. 29, 2017-Nov. 1 2017, pp. 1-3, doi: 10.1109/ICSENS.2017.8234060. (Role: offline sensor data analysis, manuscript edit/revision)
- L. Abraham, A. Urru, M. **P. Wilk**, S. Tedesco, and B. O'Flynn, "3D ranging and tracking using lensless smart sensors," in **SSI 2017**: International Conference and Exhibition on Integration Issues of Miniaturized Systems, 2017: Verlag Wissenschaftliche Scripten. (Role: experimental work and data collection, manuscript revision)

2 State-of-the-art

M. P. Wilk, M. Walsh, and B. O'Flynn, "Human motion tracking technology for healthier and longer living", Frontiers, 2020, (under review)

2.1 Technology for Human Motion Tracking and 3D Pose Detection

Human motion tracking systems play an important role in many application spaces, including motion capture, sports, fitness, rehabilitation to name a few. It is a term that describes the process of detecting and tracking the motion of human body over time. 3-D pose detection is one of the main tasks in this process. It involves determining the position and orientation of an object in 3-D space; also referred to as the 6-Degree-Of-Freedom (6-DOF) pose [19].

The interest in human motion goes back far in human history. In fact, it dates back as far as the earliest recorded history [20]. The first records can be dated back to ancient Egypt and Mesopotamia. Scientific community tends to associate the beginning of written history with the ancient Greeks and the records of their work. Aristotle's (384-322 BC) writings left the first evidence of humans interest in motion tracking. In his book 'De Motu Animalium' ('On the Movement of Animals'), he described animals as mechanical systems [21]. His works were succeeded by numerous prominent figures, including Leonardo da Vinci (1452-1519) and Galileo Galilei (1564-1643), who made some of the first attempts at mathematical modelling of human motion. Borelli (1608-1679), who is often considered the father of biomechanics, wrote a book on 'De Motu Animalium' that was published in 1680. This book shows the evidence of an understanding of the forces needed for an equilibrium in various joints of the human body. It was published well before Newton (1745-1810) published his laws on motion [22]. Figure 1 shows some of Borelli's diagrams, which show the relationship between force, mass and acceleration.

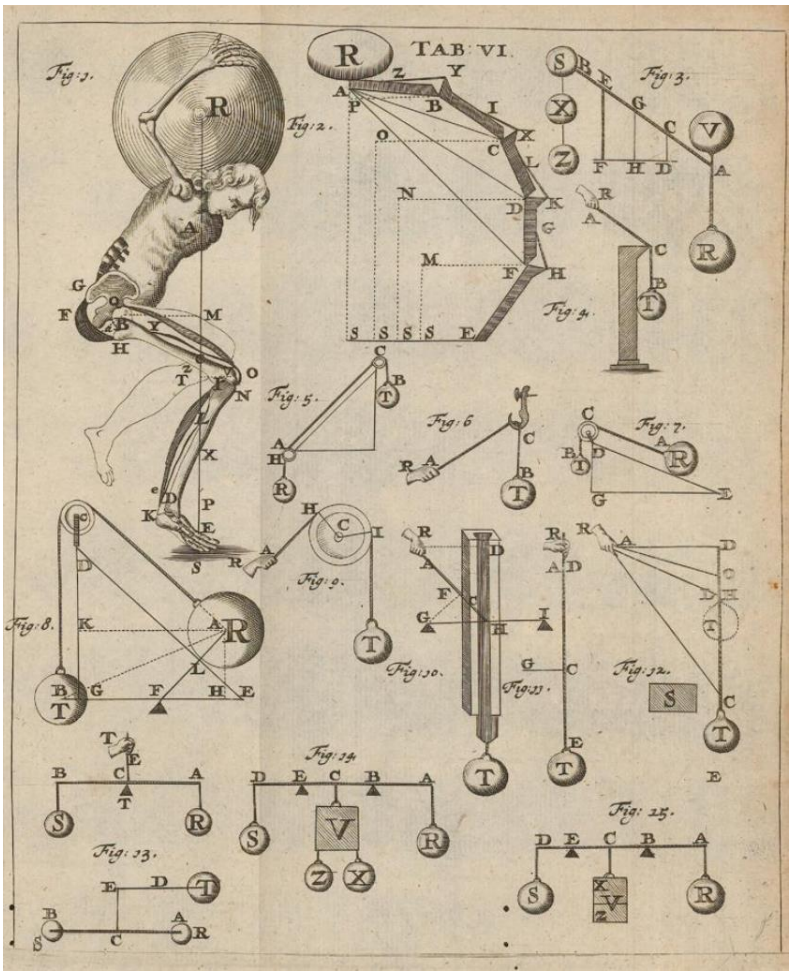


Figure 1: 'De Motu Animalium', Borelli 1680 (reproduced from [23])

The scientific advances in motion tracking continued at an increasing rate until the 19th century, mainly due to the lack of appropriate tools to comprehensively track and analyse human motion. The advent of photography was one of the first major developments that helped advance the research in human motion tracking. In the late 19th century, Edward Muybridge used multiple cameras to capture human motion in sequences of images, which was published in 1907 [4]. This was one of the first examples of having the ability to capture and analyse human motion over time, as shown in Figure 2.

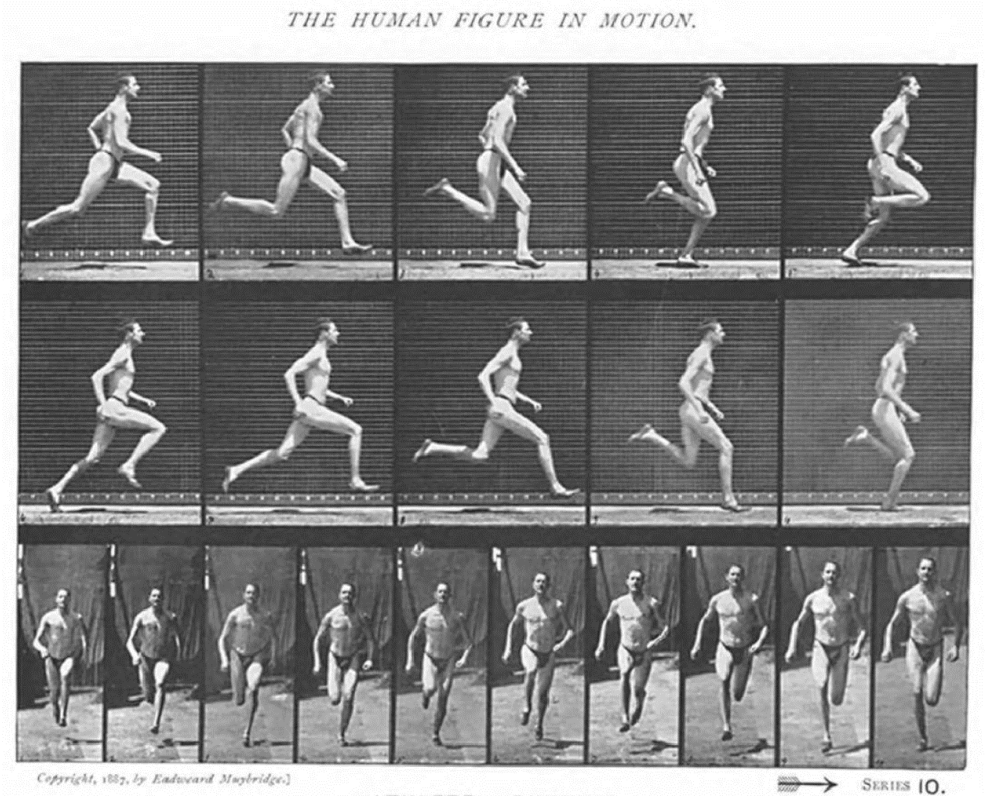
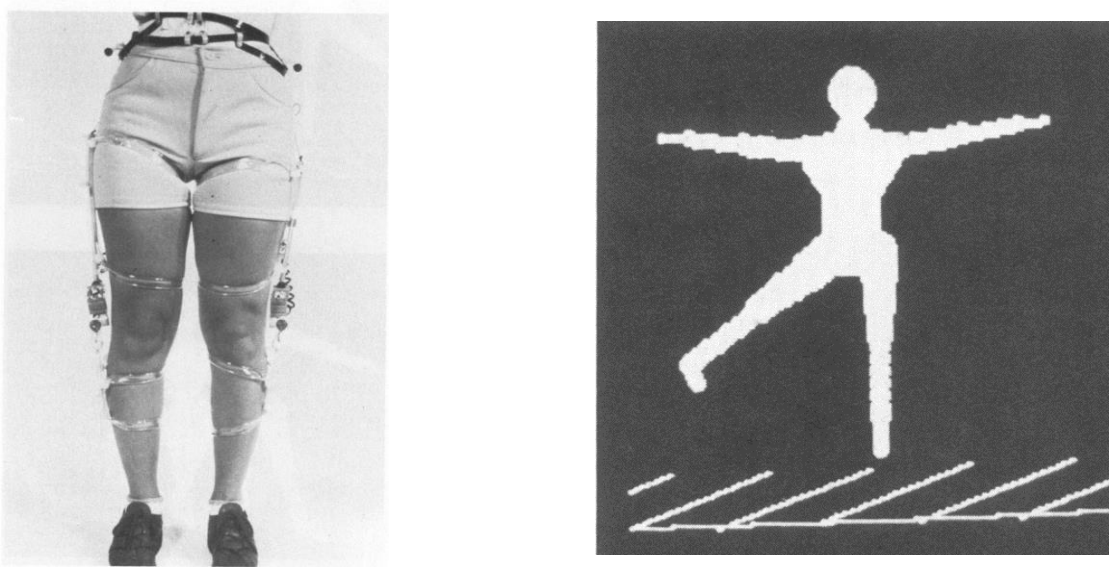


Figure 2: The Human Figure In Motion, Muybridge, 1907 (Reproduced from [4])

At that point, the use of technology became increasingly more feasible for human motion tracking. Some of the first examples of the use of human motion tracking in a practical application can be found in Disney's film Snow White and the Seven Dwarfs from 1937 [24]. The motion of a human body was tracked and used to help make the animated characters' motion more human-like; using a method known as rotoscoping, which is a technique of drawing shapes on a film, frame by frame, to create the effect of motion when played at a higher frame rate.

The current revolution in motion capture and tracking began in the 1970's. By this time, computer processing capabilities and sensor technology had matured sufficiently to be used effectively in human motion tracking applications. This was the turning point for such technology to be widely, and increasingly, used in the field of motion tracking. In 1982, researchers at Simon Fraser University began to analyze human motion using computers and electro-goniometer sensors attached to the human body. In this approach, joint flexion was tracked and used as an input into a motion animation system that rendered the scenes on an Apple II computer [25], as shown in Figure 3.



(a) (b)

Figure 3: Electrogoniometers (a) used as input to motion animation system rendered on the Apple II computer (b) (adapted from [25])

Some of the other historically notable developments in monitoring human motion include the introduction of the first image-based systems developed in the 1980s. In 1983, a stereoscopic vision system developed in conjunction with multiple LEDs were used at Massachusetts Institute of Technology to track human body motion to control stick figures [26]. The technological developments in this field have continued to advance to this day.

At present, human motion tracking is carried out using a variety of technologies, each with their advantages and disadvantages, as shown in Table 1. A more detailed description of motion trackers that use these sensor modalities is provided in the subsequent sections.

TABLE 1: MOTION TRACKERS CLASSIFIED BY SENSOR MODALITY: MAIN ADVANTAGES AND DISADVANTAGES

Sensor Modality	Strengths	Weaknesses
Mechanical	High accuracy	Limited flexibility
Acoustic	Ease	Susceptible to interferences
Radio Frequency	No need for line-of-sight	High cost
Magnetic	High accuracy	Short distance
IMU	No need for line-of-sight	Poor positional tracking
Vision	Positional tracking	Sensitive to lighting conditions and occlusions

The choice of technology and performance specifications depend on the application requirements of the end use scenario in question. Therefore, the complexity of the motion tracking system may

vary; from relatively simplistic, with inaccurate measurements, to sophisticated, highly accurate systems. An example of a relatively simple motion tracking system may be a knee flexion and extension angle measurement system incorporating an electro-goniometer; commonly used in physiotherapy [27]. On the opposite side of the complexity continuum are the advanced, infrastructure-heavy, and high-performance systems. These systems are used for motion capture in demanding applications, such as character animation in the movie industry. These systems can precisely track every major part of the human body in real time, such as Vicon or Optitrack systems, which involve multiple cameras and advanced data processing [7, 28-30]. However, high performance of such systems comes at a price. Such systems are complicated to set up, require a significant amount of expert human resources, and are costly.

Human motion tracking technology consists of two main aspects, i.e. the Hardware (HW) and Software (SW). The HW consists of various modularized building blocks of equipment, as shown on the example of the smart sensor in Figure 4. All of the elements are important including the external support infrastructure, such as the networking, telecommunications and data processing aspects. The SW plays an equally important role, as it acts as the brain that controls the operation of the HW. The sensor technology is one of the key elements in the motion tracking systems, as it largely determines the capabilities of the system. Thus, the SOA in motion tracking systems can be categorised by the sensor type, or more specifically the sensor modality of the tracker. The major sensor modalities include: mechanical, acoustic, Radio Frequency (RF), magnetic, inertial, and visual. Also, motion trackers can be divided into two broad classes; unimodal and multimodal. Whereas, unimodal trackers use a single sensor modality, the multimodal approaches combine more than one sensor modality in a single tracking system to complement the weaknesses of individual modalities.

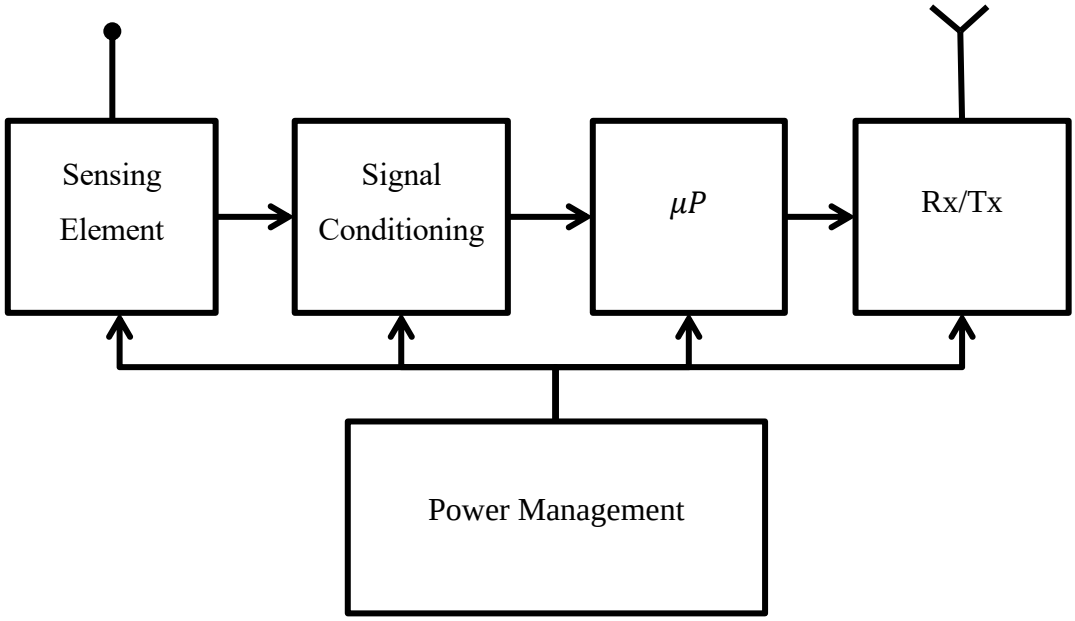


Figure 4: Generalised Block Diagram of a Smart Sensor

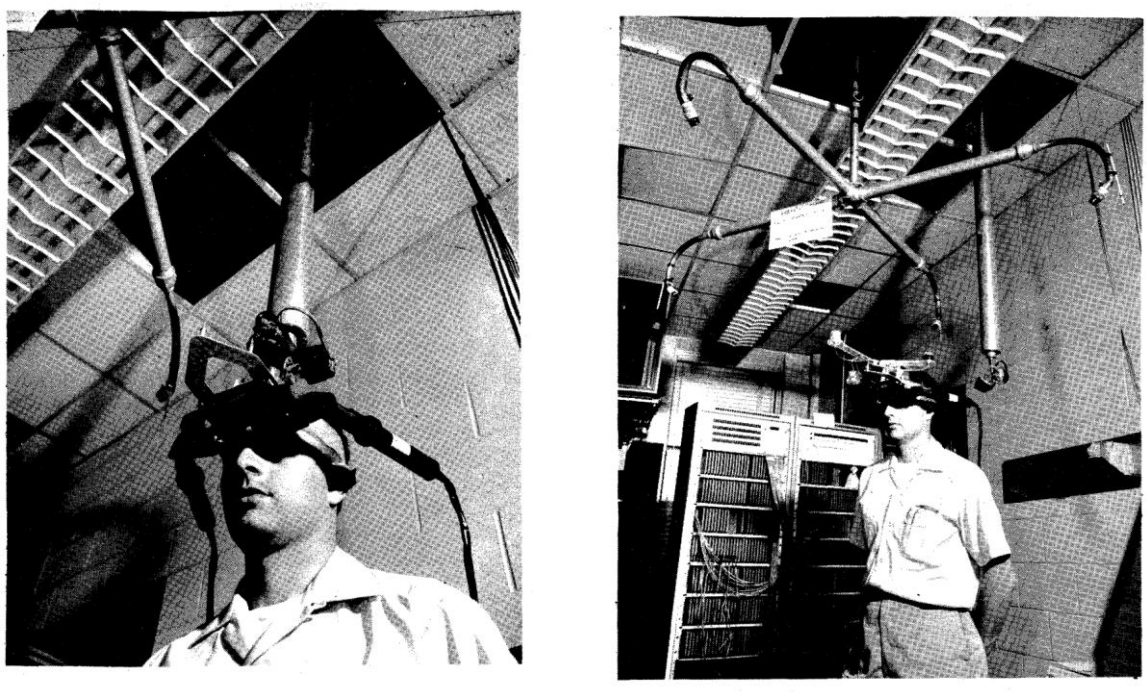
2.1.1 *Unimodal Approach*

A Unimodal system consists of a single sensor modality as the input for the motion tracking calculations. The motion tracking function is performed using a single type of information. Unimodal trackers may use one or more sensors with the same modality. For example, a stereoscopic vision system may be considered unimodal if both cameras sense the same wavelengths of light [31]. The advantages of unimodal trackers are in that they tend to be complex and therefore tend to be less expensive. Their main disadvantage is in that they cannot be effectively used in applications where a single sensor modality does not provide a sufficient amount of information to support the required analysis to the required levels of accuracy.

2.1.1.1 Mechanical Trackers

Mechanical motion tracking systems usually take the form of mechanical linkages attached to the human body parts that need to be tracked, as shown in Figure 3 (a). Those linkages are coupled with sensors, such as potentiometers. A classic example of a mechanical tracker is the head motion tracking system of Ivan Sutherland's pioneering Virtual Reality (VR) headset [5], as shown in Figure 5 (a). Flexible strain-gauge- or fibre-optic-based sensors can help replace the rigid linkages in such mechanical trackers [32]. Mechanical trackers have many advantages. The main advantage is that they can be very accurate and are not susceptible to errors. On the other

hand, they can be bulky and impractical for many applications, such as those involving whole-body tracking. Also, in order to ensure high accuracy, they need to be mounted correctly to ensure the soft tissue does not cause errors, which is a common problem among most motion tracking technologies.



(a) (b)

Figure 5: Mechanical (a) and Acoustic (b) Motion Trackers used in Ivan Sutherland's VR system (adapted from [5])

2.1.1.2 Acoustic Trackers

Acoustic trackers usually utilise multiple active ultrasound sensors, which consist of an ultrasonic transmitter and receiver pair. The transmitter generates a short pulse of signal that is detected by the receiver. Depending on the specific design choices, the time-of-flight (TOF) and/or the signal strength of the received signal are used to determine the distance between the transmitter and the receiver. A single transmitter-receiver pair can determine the 1-Dimensional (1-D) position, i.e. the distance along a straight line, while multiple sensors can be used to calculate the 3-D position, and possibly the orientation of the object being tracked. The VR headset system of Ivan Sutherland, for example, used three ultrasonic transmitters mounted on the head and four receivers installed around the head [5], as shown in Figure 5 (b). Acoustic trackers are effective at position tracking, and easier to use than the mechanical trackers. However, they also suffer from certain problems. Their main drawbacks are in the requirement of maintaining the Line-Of-

Sight (LOS). They also suffer from limited efficiency, which is associated with the size of the transducers. Their range is also limited due to the ultrasonic frequencies [33].

2.1.1.3 Radio Frequency Trackers

Radio Frequency trackers are widely used in localisation and positioning applications. In some respects, they work in a similar way to acoustic trackers, with the exception that radio frequency signals are used as the localisation technique rather than sound waves. In principle, RF trackers determine the position based on the TOF of the signal between the transmitter and the receiver. As in thecase of acoustic trackers, an RF tracking system comprises of the tracked devices and multiple points of reference, often referred to as anchor points. One of the main advantages of using RF signals is in that they can travel through various media, such as walls or smoke. It reduces the impact of LOS between the transmitter and receiver, as compared to the acoustic trackers.

There exist various RF-based tracking technologies in the SOA. One of the most relevant ones in the context of human motion tracking is Ultra-Wideband (UWB) technology, which can achieve accuracy typically below 10 cm [34]. However, the cost of RF motion tracking is also generally high. There also exist radar-based systems in the SOA that use such frequency bands that penetrate certain media but reflect off the human body. The transmitting antenna generates a signal that reflects off the human body and returns to the receiver's antenna. Adib et al. proposed a tracker based on this concept [35]. It was able to locate and track the location of the centre of human body including certain gestures with and without LOS; through a wall. However, it had the limitation of low resolution, which is typical of radar-based systems. Also, although while certainly a promising technology, it raises ethical and security concerns, such as privacy. Some of these trackers have the ability to penetrate clothing, thus infringing on people's privacy [36].

2.1.1.4 Magnetic Trackers

Magnetic motion tracker technology is based on using electromagnetic transmitters and passive, coil-based, receivers to measure the strength of a received signal. The strength of the received electromagnetic signal is related to the distance between the transmitter and receiver. In most applications, multiple transmitters are deployed to achieve a 3-D tracking of the sensor. Magnetic trackers come in a number of variants, such as time multiplexing, frequency multiplexing techniques, or transmitter coil design [37, 38]. Magnetic trackers have several advantages, such as the insensitivity to occlusions, including the human body, and high accuracy [39]. However,

they suffer from certain drawbacks. The main limitation is the relatively limited range; due to the rapid decrease of the electromagnetic signal's strength with distance. Hence, the existing solutions in the SOA tend to operate over relatively short distances [40, 41]. Moreover, the presence of metallic, ferromagnetic, objects can cause distortions to the signal.

2.1.1.5 Inertial Sensor Technology

Inertial sensor technology is one of the most popular approaches used in human motion tracking applications using wearable smart sensor systems. It is based on Micro-Electro-Mechanical Systems (MEMS) that can package what previously were large and complex scientific instruments in small Integrated Circuit (IC) chips [42]. Whereas the ICs primarily focus on the electrical properties of materials to make electronic systems, the MEMS combine both the electrical and mechanical properties of materials. MEMS are used to make various sensors and actuators, such as pressure sensors, inertial sensors, pumps or motors to name but a few [43].

MEMS accelerometers and gyroscopes are some of the most widely used sensor modalities in the context of human motion tracking. In the past, each sensor modality was made in a separate IC unit, thus forcing system designers to make larger wearable smart sensor systems. In recent years, the SOA has moved towards systems that incorporate multiple MEMS inertial and other sensor modalities in a single, highly miniaturised IC package, referred to as the an IMU [44]. A typical IMU tends to incorporate an accelerometer, gyroscope and magnetometer, such as the MPU9250 made by TDK InvenSense in a 3x3x1 mm package [45], shown in Figure 6 (a). These three modalities are combined to complement the shortcomings of the individual sensor modalities, such as drift, bias offset, and susceptibility to magnetic disturbances or general measurement errors. The data fusion algorithms use this multimodal data and accurately compute the IMU's orientation in 3-D space, such as that proposed by Madgwick et al. [46] or Mahony [47].

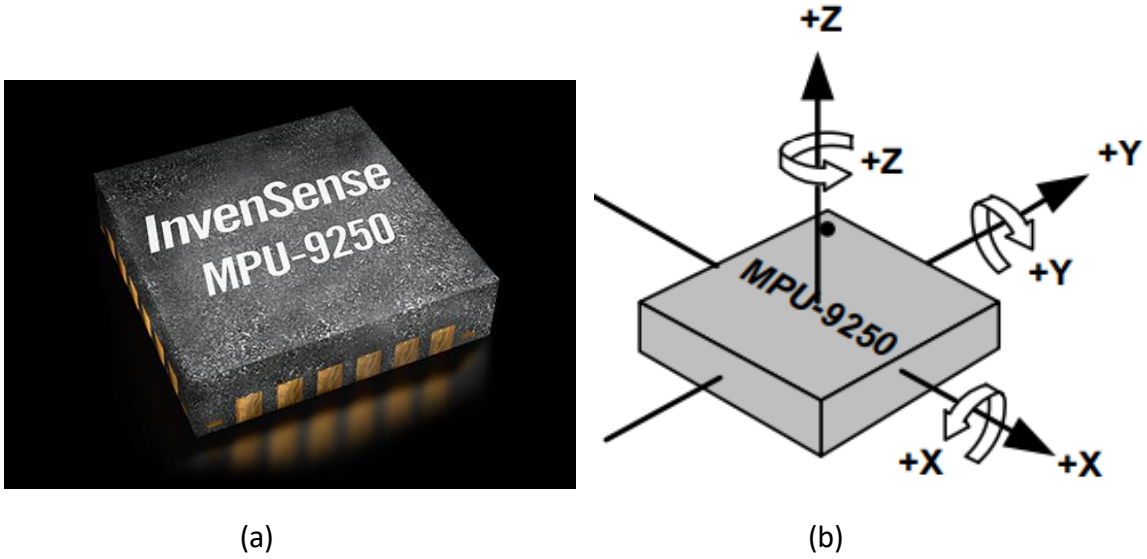


Figure 6: TDK Invensense MPU9250 IMU(a) and the Orientation Axes for Accelerometer and Gyroscope (adapted from [45, 48])

IMU sensors offer many advantages over the alternative sensor technologies. Firstly, they are miniaturised and highly energy efficient which makes them suitable for low power wearable applications. Secondly, they measure motion relative to gravity and Earth's magnetic field. Thus, they do not require any external infrastructure. Finally, they have high sensing frequency, i.e. sensor readings are updated hundreds of times per second.

The main disadvantage of IMUs in the context of human motion tracking is their inability to precisely track position over extended periods of time. The relative position can be determined only over short periods of time, by double-integrating the accelerometer readings. The position inevitably drifts over time unless an absolute point of reference is provided. As well as that, IMUs are vulnerable to magnetic field disturbances despite the advanced sensor fusion algorithms.

2.1.1.6 Vision Sensor Technology

Vision is an important sensor modality in computer systems. It performs a function that is similar to that of human eye. Human eye captures light from the surrounding environment, which contains information about it, which the brain can interpret. Likewise, computer systems use vision sensors to capture information about the environment, which is extracted and processed

by the processor. This ability makes this sensor modality suitable for human motion tracking applications.

Although there are various types of vision sensors, the most widely used image sensors are based on the Complementary Metal-Oxide Semiconductor (CMOS) technology. The low cost of manufacturing and their flexibility are some of the key contributing factors to the wide use of this technology in this context. CMOS sensors can form pixel arrays of varying resolutions, pixel sizes, or bit depths. These can also be fitted with custom lenses and optical filters to meet the specifications of the given application. Furthermore, their range of sensitivity to wavelengths, typically between 400 nm and 1000 nm, encompasses that of the human eye, which makes them ideal for sensing applications that replicate the functions of the human vision system. A typical CMOS pixel array is shown in Figure 7 (a).

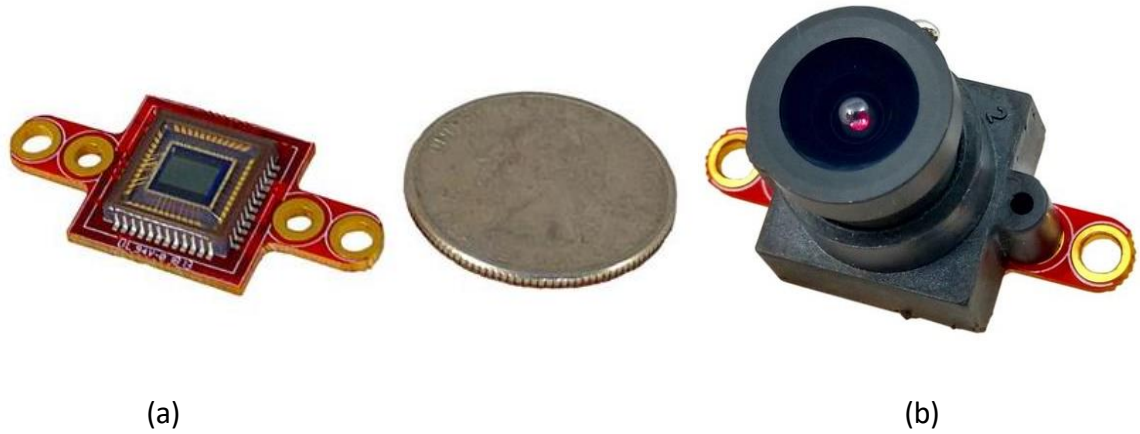


Figure 7: Active-Pixel CMOS Image Sensor MT9V034: (a) without lens, (b) with lens (adapted from [49])

In spite of its numerous advantages, i.e. the ability to operate like a human eye, vision sensors have certain limitations. On the one hand, vision systems have high computational requirements. For example, the image sensor MT9V034, shown in Figure 7 (a), has a 752 x 480 pixel array, which means that the processing unit must process 360960 pixels in each frames. Moreover, most human motion tracking applications require tens of frames per second to be captured and processed. Whereas this is not a complicated task in offline or PC-based applications, it is a significant challenge in the context of low-power, highly miniaturised, wearable smart sensor-based motion tracking applications. Secondly, the physical size, i.e. form factor has historically been a limiting factor for such applications. For instance, the lens is usually the largest component in camera modules, as shown in Figure 7 (b). Vision as a sensor modality has also other

weaknesses. One of the greatest disadvantages of vision sensors is their vulnerability to occlusions, which can adversely affect their ability to track objects in their Field-of-View (FoV). In order for the system to reliably perform the tracking, the given point of interest, i.e. feature in the images to be tracked, must remain in the camera's FoV. Secondly, cameras suffer from the correspondence problem, which is related to occlusions. As a result, the image processing algorithms can lose track of which point of interest is which in the presence of intermittently occurring occlusions. Finally, the lightning conditions have a major impact on the performance of vision systems. The less controlled the ambient lightning conditions, the more difficult it is for the system to reliably perform its intended function; and thus the more complicated the system is required to be, with regard to both HW and SW. Nevertheless, vision sensors offer many advantages that outweigh the disadvantages. The main advantage, as compared to other sensor modalities, is that they can be used to detect and track the absolute locations of multiple points of interest with high accuracy in their FoV.

Given the various strengths and weaknesses of the vision sensor modality, it has been historically used for human motion tracking carried out using expensive and infrastructure-heavy systems. Some of the most common approaches involve using multiple cameras to track multiple markers attached to the human body [7, 28]. Multiple cameras help overcome the problem of occlusions and determine the 3-D position of the given marker. The markers are normally classed as active, e.g.IR LEDs, or passive, e.g. retroreflective materials. In recent years, the introduction of RGB-D cameras, such as the Microsoft Kinect, enabled marker-less motion capture at a reduced cost and complexity [8, 9, 50].

2.1.1.7 Wearable Visual Tracking

The continuous advances in the SOA in vision sensor technology make it more feasible to consider using vision in the context of wearable miniature smart sensor systems for human motion tracking. Apart from the increasing computing power accompanied by simultaneous miniaturisation of processing units, the vision sensors themselves shrink in size. Recent works in the literature show that the lens, usually the largest component in the camera, Figure 7 (b), can by reduced in size to such an extent that the camera may be considered to be effectively lens-less [15, 51]. Although, these emerging technologies may not allow for a high quality image acquisition as compared to a regular high-resolution cameras, it is certainly sufficient to extract the necessary features, such as active markers, to perform point tracking [16]. Abraham et. al.

showed that the 3-D pose of a human hand can be tracked using two lens-less sensors in stereoscopic configuration, by tracking multiple active markers attached to it [52].

The 3-D pose of an object can also be found using a single camera in a monocular configuration. It is less demanding in terms of the hardware, as only one camera is needed. However, it comes at the expense of increased computational requirements. The monocular camera system can track multiple points of interest attached to the given object and determine its relative 3-D pose; by solving the PnP problem. The term Perspective-n-Point (PnP) was first used by Fishler et al. to describe the process of determining the pose of the calibrated camera from n correspondences between 3-D reference points, present in camera's FoV, and their 2-Dimensional (2-D) projections on the pixel array plane of the camera [53]. It is widely used in various computer vision applications. The PnP is equally applicable to fixed camera with a moving object and a moving camera with a fixed object, as shown in Figure 8 (a) and (b), respectively. The scenario shown in Figure 8 (b) is relevant to the wearable vision methodologies, which is also referred to as inside-out tracking. Likewise, the scenario shown in Figure 8 (a) is also referred to as outside-in tracking.

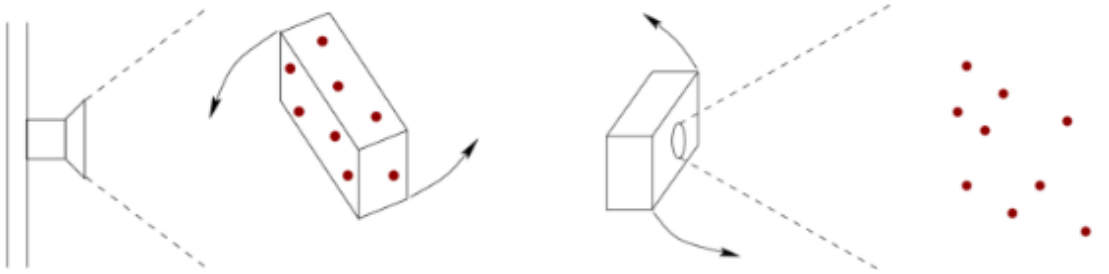


Figure 8:Perspective-n-Point Problem: (a) Fixed Camera (outside-in tracking), (b) Moving Camera (inside-out tracking)

There exist a number of algorithms in the literature that solve the PnP problem [18, 54, 55]. These methods solve it either iteratively or non-iteratively; using a varying number of reference points. The iterative methods tend to require fewer points of reference but are more computationally complex, while the non-iterative methods require more points. However, regardless of the approach, at least three points of reference are required to find the 3-D pose. Figure 9 shows how it works on the example of the Oculus Rift headset. The Oculus sensor is a fixed camera with an

IR filter attached to it, Figure 9 (a), while the headset contains multiple IR LEDs, as shown in Figure 9 (b).

(a) (b)

Figure 9: Oculus Rift: (a) Camera, (b) Headset (adapted from [56])

The use of a PnP algorithm to estimate the pose of a moving camera was shown by Oliver Kreylos using the Nintendo's Wii Remote controller back in 2008 [57]. It was later reproduced using the OpenCV open-source library. The system used four IR LED-based points of reference and the camera in the controller to track the 3-D pose [58]. This solution used the iterative PnP method based on the Levenberg-Marquardt (LM) algorithm [59, 60], as shown in Figure 10.

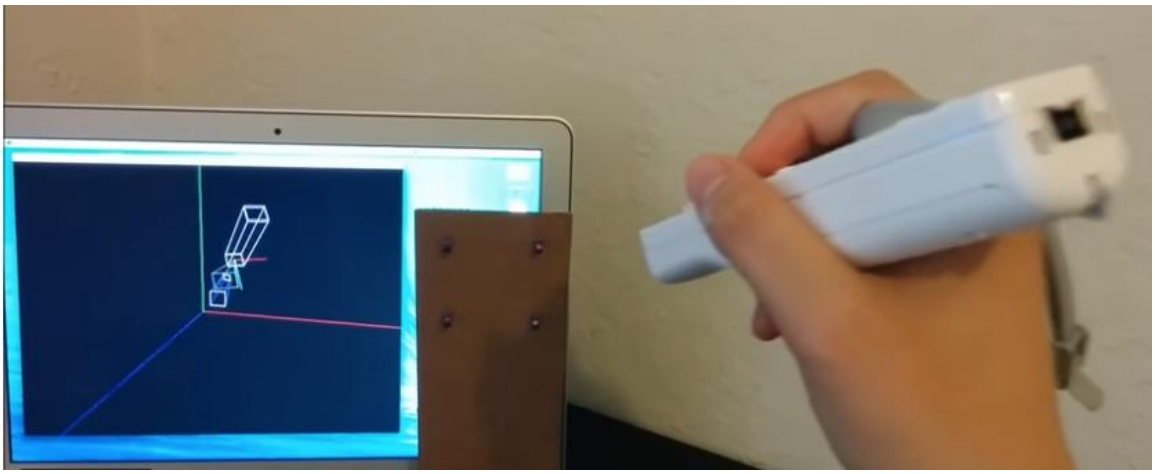


Figure 10: Pose Estimation of Moving Nintendo Wii Remote by solving the PnP problem (adapted from [58])

The LM algorithm estimates the pose by minimising the reprojection error, i.e. the difference between the predicted and measured positions of the reference points on the pixel array, given the calculated camera's extrinsic parameter matrix. Reprojection error is a metric commonly used in camera calibration procedures to quantify its quality [61], as shown in Figure 11.

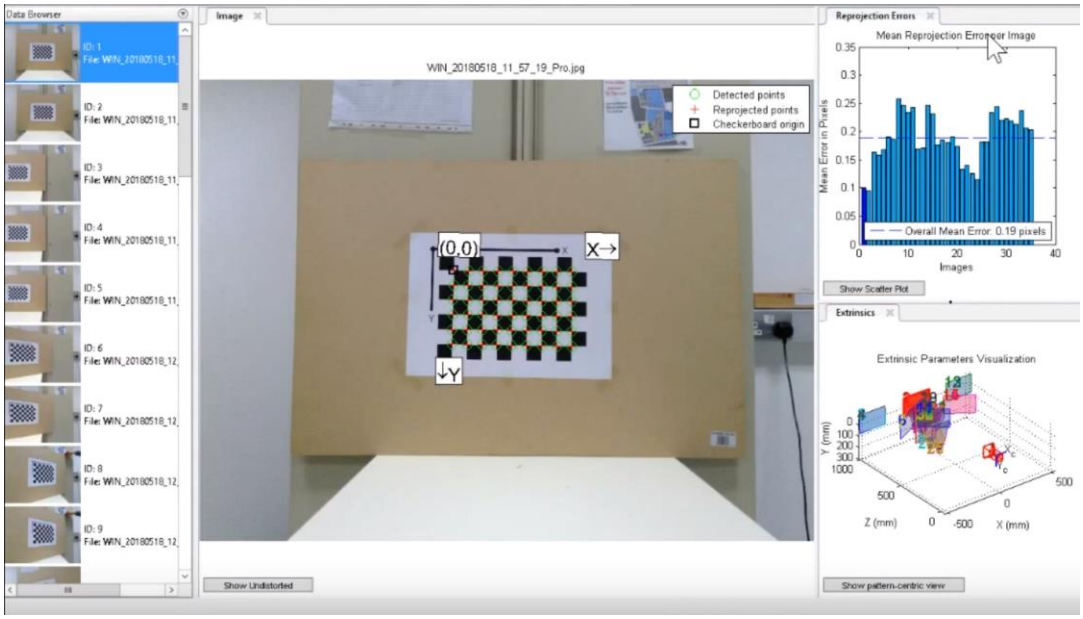


Figure 11: Reprojection Errors During Camera Calibration in MATLAB

The PnP problem is usually applied to known environments, i.e. those with a known pattern of points of interests. On the other hand, a class of algorithms exists in the literature, which is used for tracking in unknown environments, i.e. where the points of interest are unknown. It is generally referred to as the Simultaneous Localisation and Mapping (SLAM) method [17]. SLAM methods aim at mapping the unknown environment and tracking the position of the moving camera in 3-D space relative to those points. The Google Tango tablet is an example of a mobile device capable of performing the SLAM in real time [62].

The two approaches, i.e. PnP and SLAM, are able to robustly determine the 6-DOF pose of the moving camera. However, they have high computational requirements, which limit their suitability for human motion tracking using highly miniaturised, low-power, wearable smart sensor systems. These requirements are still too high for the applications space this work targets which requires mobile, resource-constrained processing capability.

2.1.2 *Multimodal Approach*

Although the motion tracking systems based on a single sensor modality, described in Section 2.1.1, can be used to perform motion tracking, they do not provide a sufficient amount of information to carry out a reliable and robust performance 6-DOF pose estimation. Their shortcomings cannot be readily overcome in a unimodal approach. This is particularly true in the context of the highly constrained, low power, miniaturised wearable smart sensor systems upon which this work is focused. Therefore, a multimodal sensor data fusion is considered in this thesis as an alternative approach.

Multi-sensor data fusion is a broad and multidisciplinary research area. It focuses on combining information from multiple sources to obtain a more complete picture of a given situation [63]. When data from a single source provides either insufficient or inadequate information to solve a given problem, multi-sensor data fusion methodologies are often considered. This was originally developed for military applications, primarily for target tracking applications. Subsequently, it was adopted in non-military applications spaces [64]. This broad engineering discipline was formally standardised into a multi-level framework, called the JDL model [65]. This model divides the fusion process into four increasing levels of abstraction, from low-level signal processing routines to abstract result interpretation. This framework serves the research community as a set of guidelines that may be followed where appropriate.

An extension to the multi-sensor data fusion approach is the multimodal approach. Multimodal data fusion discriminates between the individual modalities of the data sources. The key property of multimodal data fusion systems is the complementarity of the data sources [1]. Multimodal sensor fusion systems use multiple sensors that sense a number of modalities that complement the shortcomings of the individual sensor modalities. It should be noted that multi-sensor fusion is not necessarily the same as the "multimodal sensor" fusion. For instance, in stereoscopic machine vision systems usually two cameras are used to add the ability of sensing in 3-D, to replicate human vision. However, a stereoscopic vision, while it is indeed a multi-sensor data fusion system, is not multimodal. An example of a multimodal sensor data fusion is shown in Figure 12, which shows multiple sensors, with varying modalities, being used for human activity recognition.

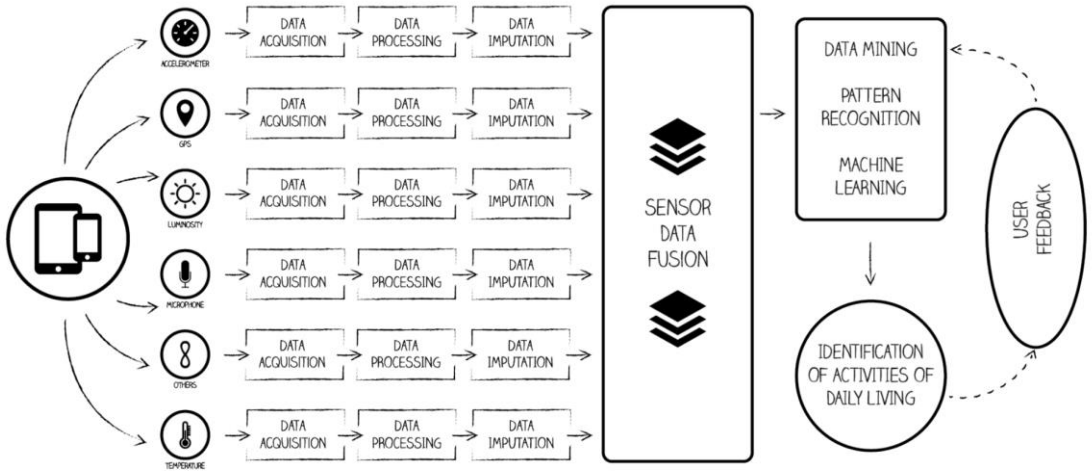


Figure 12: Example of Multi-Sensor Multimodal Data Fusion System for Human Activity Recognition (reproduced from [66])

2.1.2.1 Outside-In and Inside-Out Tracking Systems

Motion tracking systems can be broadly classed as being either outside-in or inside-out trackers. The class of the tracker depends on how the tracking function is performed. In the case of outside-in trackers, the object of interest being tracked is observed from the outside by a stationary tracking device fixed in the environment at a known position [67]. Cameras are commonly used in such tracking systems, where the cameras are mounted in the environment and capture images of the tracked objects, e.g. the human body. The motion capture systems made by Vicon Motion Systems Ltd are an example of outside-in motion trackers [29]. The inside-out trackers are the opposite of outside-in tracker systems in that they observe the environment from the perspective of the moving object itself and track its motion relative to it. In most cases, a camera is attached to the moving object and tracks points of reference in the environment to determine the object's 3-D pose. The Oculus Quest VR headset is an example of an inside-out tracker [68].

2.1.2.2 Multimodal Sensor Fusion in Human Motion Tracking

Multimodal sensor fusion is a common approach in human motion tracking applications. The complementary nature of the different sensor modalities helps solve problems that are difficult to solve otherwise. An example of such a multimodal system is one that combines vision and IMU sensor modalities to complement the weaknesses of the individual modalities. In fact, it is one of the most common approaches to human motion tracking found in the literature such as the

system proposed by Foxlin et al. [69]. The main weakness of IMU, i.e. the difficulty in reliably tracking the absolute position over extended periods of time, which is the one of the strengths of vision systems. Likewise, it is relatively difficult to determine the orientation using cameras without using complex algorithms or a purpose-designed external setup. Moreover, cameras require certain conditions, such as the line-of-sight to track the points of interest. Occlusions or uncontrolled lighting conditions, even transient ones, can cause loss of tracking. The IMU based systems can complement these weaknesses. Therefore, the human motion tracking systems based on opto-inertial sensor fusion have the potential for achieving a better overall performance in the considered application space, as compared to other approaches.

The fusion of vision and IMU sensors for human motion tracking is an active research area. Atrsaei et al. used wearable IMU sensors with a stationary Kinect sensor to track the arm motion using unscented Kalman filter, as shown in Figure 13. The system achieved a reduction in orientation error of 50 %, as compared to cases that used the individual sensor modalities separately [70]. Rodrigues et al. proposed a marker-less, multimodal, system for motion capture with multiple Kinect sensors and wearable IMUs [71]. It is an outside-in tracking system, i.e. the vision system is located outside of the object being tracked. The system was used for tracking certain individual body parts, as shown in Figure 14. It offered a less expensive and simpler to set up alternative to the more expensive and generally complex, marker-based, multi-camera, motion capture systems, such as the Vicon or Optitrack systems[29, 30].

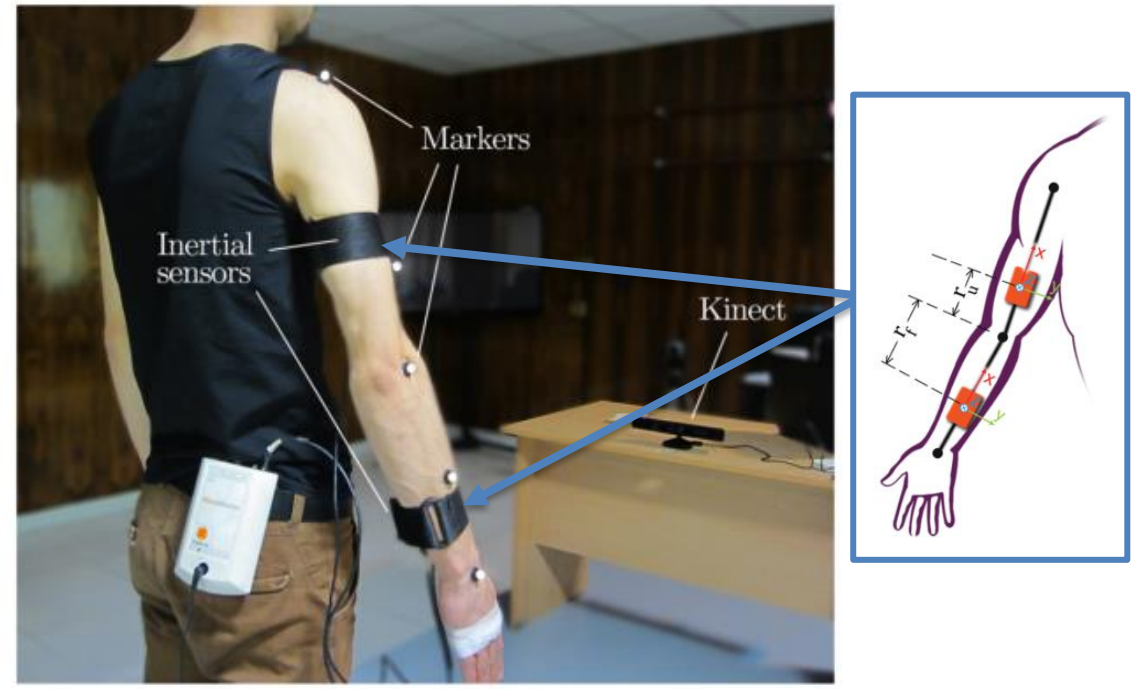


Figure 13: IMU and Vision Sensor Fusion for Arm Motion Tracking (adapted from [70])

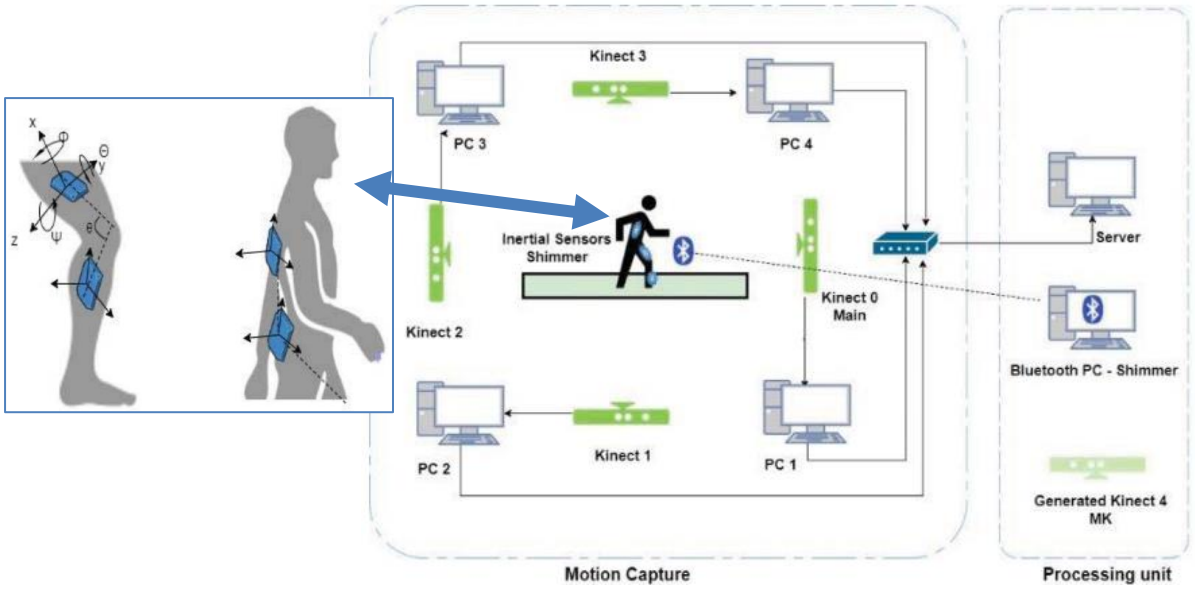


Figure 14: Multimodal Motion Capture System with Wearable IMU and Multiple Kinect Sensors (adapted from [71])

Likewise, an inside-out system that performed an the opto-inertial sensor fusion was proposed by Feng et al [72]. This system performed SLAM for motion tracking. The problems caused by jitter and low frame rate of the camera were reduced by the orientation data from the IMU. The advantage of this approach is in that SLAM does not require any pre-installed points of reference in the environment. Instead, this system automatically finds a large number of points of reference and keeps track of them from frame to frame. However, this is achieved at the expense of high computational requirements. For instance, this system required a high-end smartphone to process the input images at 30 frames per second. Therefore, the cost and size of such solutions remains prohibitive in terms of the affordable, low power, and highly miniaturised wearable smart sensor systems for human motion tracking.

Solving this challenge is an active and exciting topic in the research community. The literature shows evidence of the SOA moving towards more energy efficient increasingly miniaturised solutions. It shows advances in both development of algorithms and the associated system architecture that supports human motion tracking. A series of research projects involving Eric Foxlin et al. shows a range of evolving inside-out opto-inertial trackers that improve in performance without significantly increasing the computational complexity [69, 73-75]. The IS-1500 is the most recent product their research work has developed [11]. The IS-1500 achieved a typical accuracy of 2 mm (accuracy metric was not specified by the manufacturer) in position using four external points of reference (termed passive fiducial markers). However, in spite of the high performance, lower size, and overall system complexity, the system continues to have high computational requirements. A less complex outside-in tracker was proposed by Li et al. that combined a monocular camera and a moving object that contained two passive (fiducial marker) points of reference and an IMU [12]. This system's Root Mean Squared Error (RMSE) was below 5 cm at a distance of 113 cm. A similar low-cost system was proposed by Maereg et al. that also used an external camera that tracked a moving object, a human hand, which contained two IR LEDs and an IMU sensor [13]. This system is shown in Figure 15. The position calculation was carried out by a simplified mathematical formulation, based on proportionality between the camera and the two points of reference, complemented by orientation data from the IMU. The RMSE in position estimation was less than 0.66 cm. However, the size of the work envelope was not large (where, work envelope is defined as a volume of space within which the system can operate within its specifications), with maximum distance along the z-axis being 30 cm.

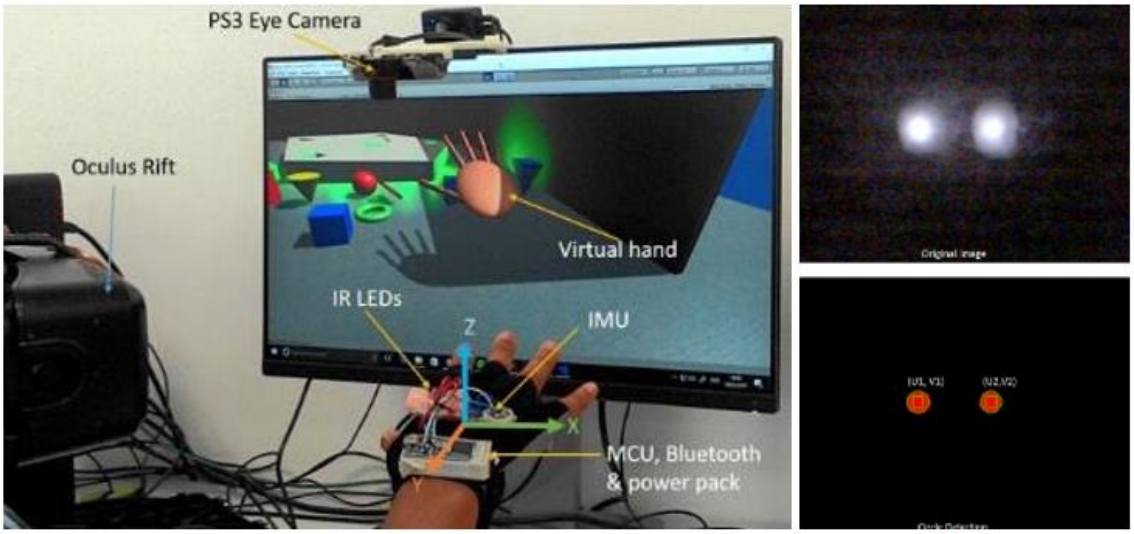


Figure 15: Outside-In Hand Tracking System (adapted from [13] ***)***

In recent months, a SOA inside-out motion tracking system for VR/AR applications has been introduced to the market. It is a commercial system called Antilatency [76]. It is an opto-inertial tracker that incorporates the IMU and vision sensor modalities in the wearable unit, as shown in Figure 16 (a) and (b). The wearable unit performs sensor fusion using the IMU and camera, which tracks active IR LED markers embedded in the floor, shown in Figure 16 (c). As to the technical information on this system, not many details were found in the scientific literature at the time of writing this document. The company Ant Inc., which sells this product, has not revealed much information about it; beyond generic details. According to the company's website, the system has a high update rate of 2000 updates per second, accuracy in position tracking at submilimeter level robustness against occlusions [76].

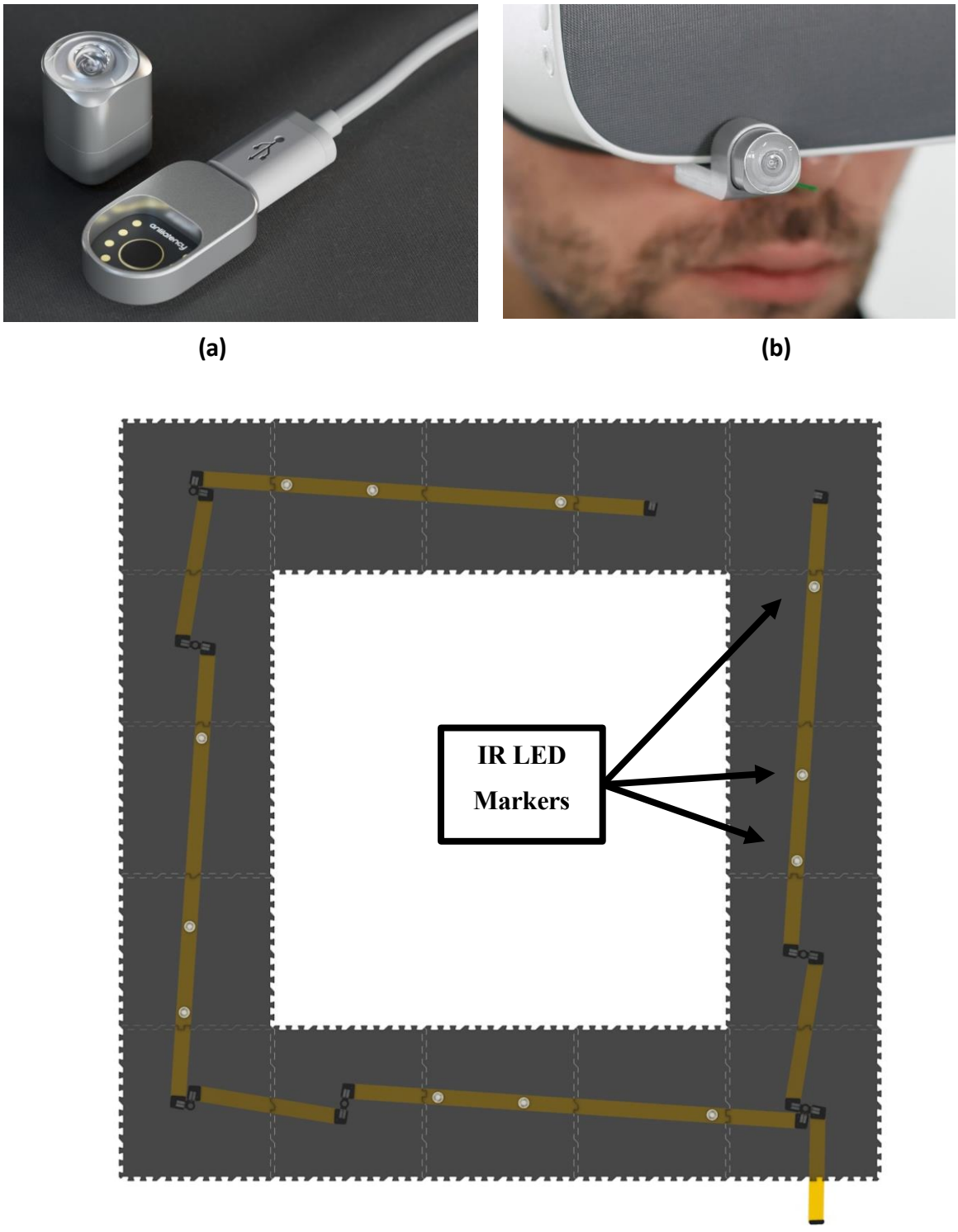


Figure 16: Antilatency Inside-Out Tracker: (a) Wearable Opto-Inertial Motion Tracker, (b) Motion Tracker Attached to VR Headset, (c) Floor Mat with Active IR LED Markers

2.2 Gap in the State-of-the-Art in Wearable Human Motion Tracking Systems

The advances in the SOA in various aspects of motion tracking technology are some of the enabling factors for low-cost, low-power and highly miniaturised wearable human motion tracking systems. While the unimodal methodologies are insufficient for achieving a reliable, accurate, and robust motion tracking performance, as outlined in Section 2.1.1, recent advances in the SOA in motion tracking using multimodal sensor fusion are make it feasible, as outlined in Section 2.1.2.

The main disadvantage of using IMUs in the context of human motion tracking is their inability to precisely track the absolute position over extended periods of time due to drifts associated several factors, such as the accumulation of error when double-integrating the acceleration measurements. Even the most advanced position tracking systems that are based on IMUs eventually lose track of the position due measurement errors. The weakness of IMU in terms of position tracking is one of the main strengths of vision-based position trackers. Cameras are commonly used in tracking applications due to their ability to capture points of reference in the environment, which can be used for detecting and tracking the absolute position of the given object of interest. However, vision-based systems are dependent of the lighting conditions and a direct LOS. The accuracy of motion trackers based on vision sensors decreases in unfavourable lighting conditions and presence of occlusions, which do not affect the IMUs. This is one of the main reasons why these two sensor modalities are commonly used together in motion tracking applications.

The IS-1500 tracker is certainly one of the leadings systems in this research area with its high accuracy in positional tracking. However, despite the performance, its overall system implementation requirements are high [11]. Some of the requirements include: a high-resolution camera, at least four specific points of reference in the ambient environments and a high processing power to perform the tracking as interactive framerates. Although, the tracker itself has a small form factor, it needs to be connected external computing unit, as shown in Figure 17.

Figure 17: IS-1500 Opto-Inertial Tracker with a USB-C Connector to a Computing Unit (adapted from [11])

On the other hand, the two outside-in tracking systems proposed by Li et al. and Maereg et al. show the potential of a full 6-DOF pose estimation with significantly less complex system architectures and algorithms, despite lower accuracy [12, 13]. Both methods rely on two points of reference as opposed to at least four in the IS-1500 specification requirements. Both systems use the IMU data to complement the shortcoming of the monocular vision sensors, such as cameras. However, both systems are outside-in trackers, which can be a limiting factor in terms of scalability due to the requirements of an increasing number of generally expensive and complex external cameras and computing capabilities to facilitate larger tracking space.

Although various aspects of multimodal, opto-inertial, sensor fusion have been thoroughly researched and extensively reported in literature, there are certain areas that continue to be unexplored. There exists a gap in the SOA with regard to incorporating these two sensor modalities in a single resource-constrained wearable unit so as to be able to decrease the cost and complexity of the human motion tracking system, while maintaining a sufficient level of tracking accuracy. The reduction of the number of points of reference has a direct impact on the computational complexity of the wearable opto-inertial motion tracking systems. To date, there is no evidence in the existing literature on inside-out, opto-inertial, motion tracking systems that compute the 3-D pose using two points of reference, which is a significant gap in the current SOA, as the number of reference points has a direct impact on the complexity and cost of such motion tracking systems. The use of an IMU in complementing a vision system that tracks two external points of reference using multimodal sensor fusion with an inside-inside-out with a wearable device in the context of low-cost has not been reported in existing literature, thus creating the gap in research that needs to be explored. A summary and comparison of the main

properties of the three motion tracking systems, including the number of required points of reference, is listed in

Table 2.

TABLE 2: COMPARISON OF MAIN PROPERTIES OF STATE-OF-THE-ART OPTO-INERTIAL MOTION TRACKERS

Method	Position Error [mm]	Markers Required	Tracking Type	Work Envelope Size (along z-axis) [m]	Overall System Complexity
IS-1500 (Pose Recovery Algorithm with Fiducial Markers) [11]	2 (Typical) (metric not specified)	At least 4 (Passive Fiducial)	Inside-Out	Variable	High
Maereg, et al. [13]	0.21 (Static) (RMSE)	2 (Active)	Outside-in	0.045	Low
Li, et al. [12]	48.3 to 275.4 (Static) (RMSE)	2 (Passive)	Outside-In	1.13 to 4.13	Low

2.3 Hypothesis of this Work

Given the gap in research that was identified and described in previous section, the hypothesis in this thesis is as follows: We consider a low-cost, low-resolution, monocular camera system that is combined with an IMU in a single miniaturised wearable smart sensor unit, and it was coupled with two stationary points of reference, using active markers such as IR LED. Then the 3-D pose, i.e. the 3-D position and orientation, of the wearable unit could be efficiently determined. This approach has not been reported in existing literature. Moreover, the orientation data from the IMU could be used to directly complement the missing pieces of information from the vision sensor, thus reducing the overall system complexity; by avoiding the need for computationally expensive algorithms for computing the 3-D pose, such as the PnP solutions. As a result, the complexity of the sensor fusion algorithm for the 3-D pose estimation can be reduced and, thus, lead to lower requirements in terms of processing power and energy consumption. These requirements can be further decreased by reducing the computational load associated with the image processing tasks when detecting points of reference in images acquired by the camera. To that end, resolution of the camera can be reduced while introducing subpixel point detection techniques to finding the coordinates of the two points in the input images. The subpixel point detection can prevent the loss of precision of point detection caused by lowering camera's

resolution. This results in a less complex and less expensive inside-out motion tracking system, as compared to the IS-1500 tracker.

3 Wearable Vision; Point Detection and Tracking

M. P. Wilk, M. Walsh, and B. O'Flynn, "Extended Efficient Sub-Pixel Point Detection Algorithm for Point Tracking with Low-Power Wearable Camera Systems," IEEE Transactions on Image Processing, 2019, (under review)

M. P. Wilk and B. O'Flynn, "Reference Point Estimation Technique for Direct Validation of Subpixel Point Detection Algorithms for Internet of Things," in 2019 30th Irish Signals and Systems Conference (ISSC), 17-18 June 2019, pp. 1-5, doi: 10.1109/ISSC.2019.8904921.

3.1 Introduction

Visual point detection is an important research topic in the field of digital image processing. The ability to precisely determine the coordinates of a given point of interest in an image is fundamental in many image processing applications. Many high-level algorithms, such as those used in object detection, pattern recognition, spatial mapping, etc., rely on the performance of the underlying lower-level algorithms, such as image segmentation and feature detection. Point detection plays an important role in such tasks [77]. With the continuous advances in miniaturised sensor technologies, new application spaces emerge. The miniaturisation of vision sensor technology is particularly encouraging. The emerging lens-less, or planar meta-lensed, vision sensors show that the dependency on traditional lenses, typically the largest components of vision sensor systems, can be eliminated. Thus, their physical size can be reduced [14, 15, 78, 79]. The advances in the state-of-the-art in the traditional, lensed, camera systems' miniaturisation are also encouraging [80]. Therefore, the addition of vision sensor technology to low-power wearable devices used for motion tracking becomes increasingly feasible. Optical sensors can improve the positional tracking of wearable devices.

Although such highly miniaturised lens-less cameras do not necessarily achieve the same performance as their traditional counterparts for image acquisition, these can still be suitable for certain applications, such as those that involve point detection and tracking. Such cameras can be particularly suitable for miniaturised wearable motion tracking applications, where: low-power, small form factor, and long battery life are important considerations. the development of wearable vision systems can be challenging in this context. Digital image processing techniques are generally computationally intensive. Whereas it is not a limiting factor in traditional image processing applications where virtually unlimited resources are available (i.e. PC based

systems), it can be, however, problematic if the image processing is carried out on miniaturized, low-power, resource-constrained wearable devices [81]. The computational complexity of an algorithm executed on a wearable device depends on several factors. Firstly, the image frames need to be processed at interactive rates; at least 25 frames per second. A sufficiently high frame rate is required to be able to faithfully reproduce the motion of the human body. Furthermore, each image frame often needs to be processed by the image processing algorithms in multiple stages; before proceeding to the analysis of the next input frame. Furthermore, the resolution of the imaging sensor has a major impact on image processing speed. Although higher resolutions can help capture more information from the environment, it occurs at the expense of either increasing the processing power requirements or decreasing the frame rate. On the other hand, low-resolution image frames can help to increase the frame rate, but the accuracy and precision of the output may be compromised.

These challenges are difficult to tackle if the system design work is limited to the wearable device only. However, a more holistic approach that considers the ambient environment as part of the system, can be beneficial. The wearable device can be coupled with the ambient environment to get a level of control over the optical sensing. For example, consider active markers as points of interest that the camera needs to track. The wearable device can control the intensity of these points of interest to ensure that they remain well above the noise floor on the pixel array, thus leading to the reduction of the computational complexity of the image processing algorithms [82]. The decreased computational complexity can be achieved by eliminating the unnecessary sources of noise. The noise floor in the image frames can be also lowered and made uniform. It can be accomplished in the following way. In a typical point tracking application, the point detection algorithm is focused on finding the coordinates of 'blobs' that represent the points in the image. The 'blobs' can be extracted from an image by making several assumptions. For example, the points to be tracked can be specific point light sources, e.g. Infrared IR LED, and the vision sensor can be fitted with a matching optical IR filter. As a result, only the point sources representing the expected points are captured by the imaging sensor. Also, the peaks of the detected points can be well above the noise floor, thus easily detectable. The intensity of the IR LEDs can be controlled to ensure that no pixels in the imaging sensor are saturated, which makes it more difficult to find the centre of the 'blob' and achieving more accurate position measurement. Moreover, the Field-of-View (FoV) of the imaging sensor can be reduced to pixels that lie within such a radius that the geometric distortions can be neglected [83]. Under these

conditions, the sources of light can appear as Point Spread Functions (PSF) with Gaussian characteristics [84], over an area approximately between 3x3 and 6x6 pixels. Such coupling between the wearable camera and the ambient environment can significantly increase the efficiency of point detection algorithms at pixel level. Indeed, the pixel-level point finding algorithms can be limited to finding the local maxima in the image. Also, the resolution of the sensor can be decreased to reduce the number of pixels to be processed in each frame, thus further increasing the speed of pixel-level point detection algorithms.

However, in most point tracking applications, a lower resolution image decreases the accuracy of point detection. This is a limitation which can be overcome by finding the coordinates of the points at sub-pixel level. The true coordinates of the points are located around the detected peaks at the pixel level. The coordinates of the points can be refined to sub-pixel level by inspecting the neighbourhood of the peak pixel intensity, thus overcoming the limitations of the pixel resolution of the imaging sensor. The pixel intensities adjacent to the peak contain the necessary information to estimate the location of the true intensity peak at sub-pixel level. Figure 18 depicts a typical point source of light with properties of a Gaussian distribution on a pixel array; sampled at pixel- and sub-pixel levels. The super-resolution methods for sub-pixel point detection are well documented in the literature [84-89]. Some of the main application spaces for such methods include microscopic imaging and astronomy, or media encoding techniques, e.g. motion compensation in MPEG-4, for instance. Historically, the ratio of the time taken by a computer program to detect a point at pixel level was much higher than the time taken to detect the point at sub-pixel level. Therefore, more attention has traditionally been paid to the accuracy of the sub-pixel detection algorithms than the time requirements of the computation as this was seen to be negligible due to high processing power of the computing platforms. This is not always the case in the context of ultra-low-power wearable platforms. The performance of such resource-constrained systems is much more dependent on the system's and algorithms' complexity. This is particularly the case for systems that rely on the intelligent coupling of the vision sensor with the ambient environment, such as that described previously. In this case, the point detection at pixel level can be simplified to such a degree that the timing of a given sub-pixel detection algorithm may become as important as its accuracy. This work considers these two criteria as equally important since the described algorithm is mainly intended for resource-constrained wearable systems.

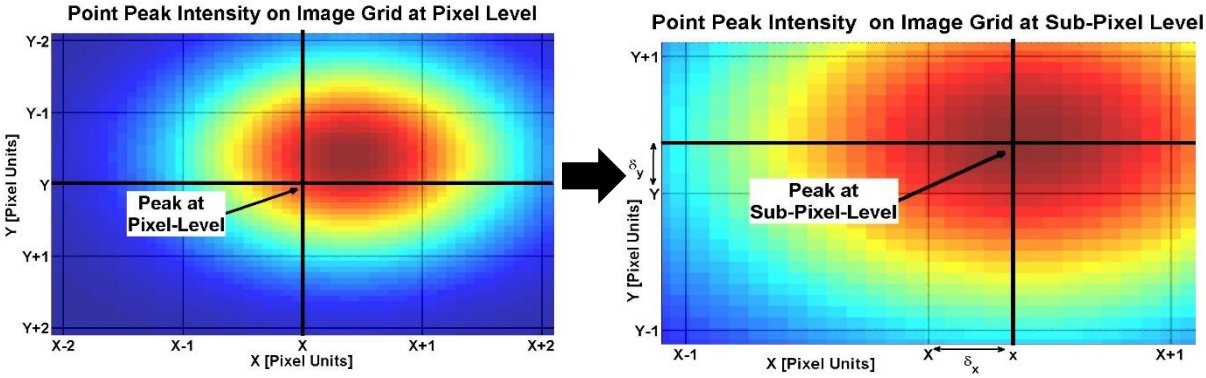


Figure 18: Gaussian Intensity Peak at Pixel and Sub-Pixel Level

3.2 Subpixel Point Detection Algorithm

3.2.1 *State-of-the-Art in Subpixel Point Detection Techniques*

Linear interpolation is a common approach for estimating the values of a function between its known discrete values. It assumes a linear relationship between the values of the function at points that surround the interpolated value. It is one of the simpler and often most efficient ways to perform the interpolation, such as that based on the 1st order Newton's Divided Difference method [90]. However, its direct application to sub-pixel peak detection is not possible. Whereas a typical interpolation problem involves finding the intensity value at a specific and known location, sub-pixel peak detection is aimed at finding the coordinates of the true intensity peak, where neither the coordinates nor the intensity of the true peak is known. The point detection algorithms can rely only on the pixel intensities adjacent to the true peak, as shown in Figure 19 (a). It shows a typical 1-D scenario with the Gaussian PSF sampled at the pixel resolution of the camera with the location of the peak refined to the sub-pixel level. The coordinates of the intensity peak at sub-pixel level are defined by x and y, as defined in (1).

$$x = X + \delta_x;\ y = Y + \delta_y \tag{1}$$

The X and Y are the pixel-level x-y coordinates, and δ_x and δ_y represent the displacements, also referred to as the sub-pixel offset, of the true peak from the detected pixel-level peak at the coordinates X-Y. The pixel-level coordinates are refined to sub-pixel level by finding the values

of δ_x and δ_y. The system can be considered as a symmetrical one, i.e. the pixel intensity profile of the point has the same properties on the image plane along both dimensions. Therefore, an analysis along a single axis is sufficient to determine performance of the system along both axes.

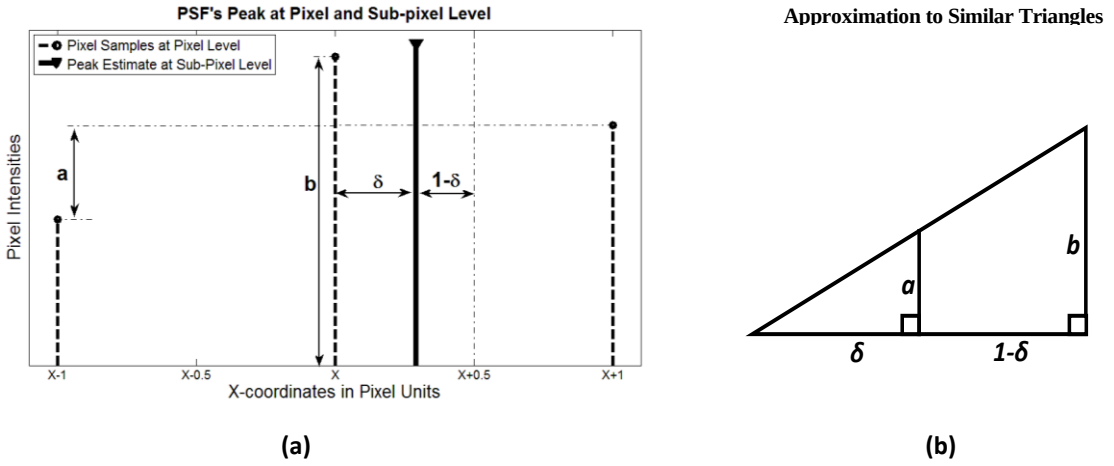


Figure 19: (a) Point Source Peak's Intensity Profile and Terms of SLI's Model; (b) Approximation of SLI's Terms to Similar Triangles

One of relevant reference algorithms covered in the literature is the Linear Interpolation (LI) algorithm, as described in [85]. Due to the assumption of linearity, it is computationally efficient when compared to other comparable algorithms. It defines δ_x as half the ratio of the difference between the preceding and the following pixel intensities, in Figure 19 (a), to the difference between the peak pixel intensity and the lower peak of the two surrounding pixels (the peak located at X-1 in Figure 19 (a)). Its accuracy is lower when compared to slower methods, such as the Gaussian Approximation (GA), [85]. The GA assumes a Gaussian spread of the intensities around the observed peak. It defines the sub-pixel offset δ_x in a similar way to that of the LI, but it differs in that it is based on a ratio of natural logarithms of the pixel intensities around the observed peak intensity.

There exist many other algorithms for super-resolution point detection in the literature. However, most of them are not suitable in the considered application space due to their high computational complexity. For this reason, the following sections focus only on the LI and GA. The LI was mathematically the closest to the proposed SLI while the GA had the highest accuracy of the three algorithms [85].

3.2.2 *Simplified Linear Interpolation Method*

Once the coordinates of the point source have been located, the sub-pixel point detection methods, such as the proposed SLI algorithms developed as part of this research, can be applied. It is computationally efficient and can find the sub-pixel offset δ_x faster than other comparable methods, as shown in Table 3. This section discusses the SLI algorithm in detail. The assumptions made in this method are similar to those of the linear interpolation, i.e. the linear relationships. However, it uses this relationship differently from the methods described in the Section 3.2.1. The underlying principles of the SLI algorithm can be explained using the trigonometric properties of similar triangles, as shown in Figure 19 (b). The pixel-level intensities of the peak and the two surrounding pixels, from Figure 19 (a), are approximated to the sides, *a,* and *b,* of the two similar triangles. Also, the sub-pixel offset from the observed pixel-level peak, *δ,* forms the horizontal side of the smaller triangle. The uncertainty area, i.e. the distance between *X* and *X±0.5* is equal to one, because this is the maximum absolute value that the sub-pixel offset δ may have around the given observed peak without having an error at pixel-level. Indeed, δ lies within ±0.5, as depicted in Figure 19 (a) and (b).

The SLI relates the pixel intensities at and around the observed peak to the sub-pixel offset δ_x as a ratio of the difference between the pixel intensities of the two pixels surrounding the observed peak to the pixel intensity of the observed peak, as in (2):

$$\delta_x = \frac{a}{b} = \frac{f(X+1) - f(X-1)}{f(X)}; \quad \delta_x \in < -0.5, 0.5 > \tag{2}$$

The maximum value of the computed δ_x is capped to $\delta_x = \pm 0.5$ pixel. Moreover, due to the way the numerator of SLI is constructed, the sign of the resultant sub-pixel offset δ_x is determined automatically.

The SLI can also be derived from the LI method, when one assumes that the system operates at the optimum operating conditions for the SLI. This relationship was found when further analysing the results of the simulations. It is briefly explained in the following section. This relationship was not obvious, until the results of the simulations of the system at the optimum conditions were analysed in detail. One of two cases of the LI is defined in (3). Although the similarity to the SLI, (2), can be observed, its numerator is more complex. However, under the optimal conditions of the system, the intensity profile of the point source's peak acquired certain

characteristics. Figure 20 shows the pixel intensity profile under optimal conditions in detail where the optimal conditions are defined as such that the standard deviation σ of the Gaussian distribution is approximately equal to 1.2. The simulated sub-pixel offset was $\delta_\mu = 0.5$. It shows that the ratio of the pixel intensities $f(X-1)$ to $f(X)$ is approximately a half and is constant. Thus, it can be shown that the denominators of the LI and SLI are approximately equal, as shown in (4). Therefore, both methods are approximately equal, as indicated in (5) and (6), under these conditions. Thus, the SLI could be considered a simplification of the LI, under these specific conditions.

$$LI = \frac{f(X+1)-f(X-1)}{2\big(f(X)-f(X-1)\big)}\ ; if\ \big(f(X+1) > f(X-1)\big) \tag{3}$$

$$\because 2\big(f(X)-f(X-1)\big) \cong f(X) \cong 0.3048 \tag{4}$$

$$\therefore \frac{f(X+1)-f(X-1)}{2\big(f(X)-f(X-1)\big)} \cong \frac{f(X+1)-f(X-1)}{f(X)} \tag{5}$$

$$\therefore LI \cong SLI \tag{6}$$

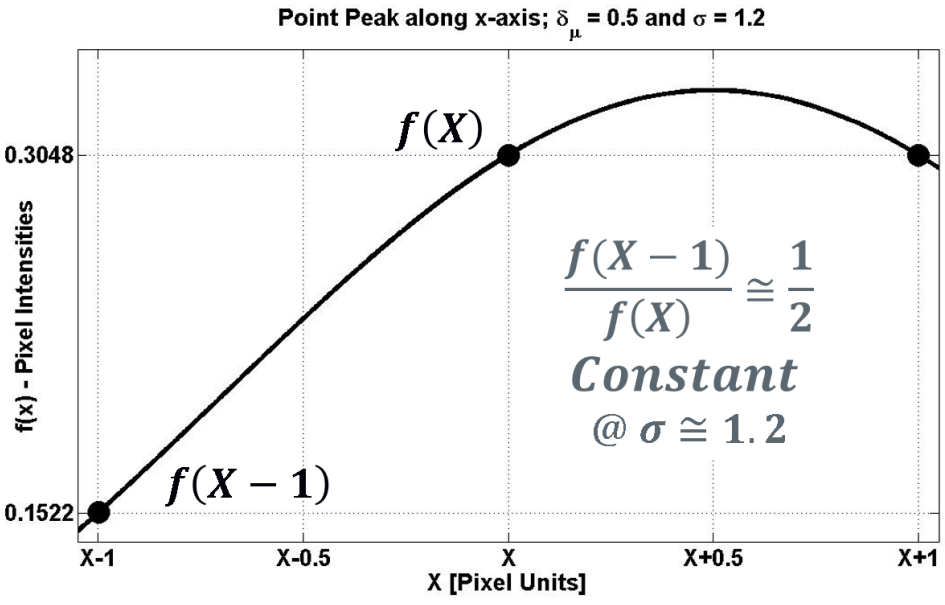


Figure 20: Simulated Pixel Intensity Profile under Optimum Conditions for the SLI

Given the mathematical relationship in (5), it can be observed how the SLI has a lower computational complexity than the LI, hence higher execution speed. Furthermore, the simulations also showed that the SLI was expected to achieve a comparable or better accuracy. The experimental validation of these expectations is the main objective of the subsequent sections of this work.

3.3 Performance Simulation

This section briefly summarizes the simulations that were carried out to evaluate the expected performance of the SLI [91]. The SLI was compared to the LI and GA with two main criteria; RMSE and the Relative Mean Execution Time. The simulations were based on the Monte Carlo approach, and the assumption of a Gaussian distribution of the pixel intensities around the peak. These results were used to analyse and predict the behaviour of the SLI under the experimental test conditions.

The intensity peak of the point source in the image was simulated by generating a 1D Gaussian distribution, along x-axis, with two input parameters $f(x|\mu,\sigma)$, where μ is the mean of the distribution μ and σ is the standard deviation. The value of μ was associated with the known sub-pixel level reference offset δ_μ through the following relationship,:

$$\mu = X + \delta_\mu \tag{7}$$

The two input parameters of the Gaussian distribution were bounded to $\delta_\mu \in < 0,0.5 >$, $\sigma \in < 0.5,3 >$. The values of the generated Gaussian distribution $f(x|\mu,\sigma)$ were used to sample the simulated pixel intensities, at integer values of x, and used in computing the sub-pixel offset δ_x using: SLI, LI and GA methods. The computed offsets were used to compute the error, as shown in (8):

$$error = \delta_\mu - \delta_x \tag{8}$$

Several test scenarios were constructed by manipulating this input parameter pair $< \mu, \sigma >$. In each simulation run, an $N = 10^6$ samples of input parameter pairs are generated for each

simulation scenario. The comparison criteria for the simulations are defined in equations (9) and (10).

$$RMSE = \sqrt{\frac{\sum(\delta_{\mu} - \delta_x)^2}{N}} \tag{9}$$

$$Relative\ Mean\ Execution\ Time = \frac{\sum_{i=1}^{N} t_i}{N} \tag{10}$$

In the simulations, the SLI was evaluated under several scenarios. In the context of the extended work, Scenario 1 was the most relevant one. In this scenario both input parameters to the Gaussian distribution function $< \mu, \sigma >$ were random, while remaining within the specified bounds. The results are shown in Table 3. The SLI achieved the highest RMSE, and lowest execution time.

TABLE 3: SIMULATED RESULTS: SCENARIO 1; RANDOM <M, Σ>

SLI		**LI**		**GA**	
RMSE $[\boldsymbol{pixel}]$	***Time*** $[\boldsymbol{\mu s}]$	***RMSE*** $[\boldsymbol{pixel}]$	***Time*** [μ***s***]	***RMSE*** $[\boldsymbol{pixel}]$	***Time*** [μ***s***]
0.1462	0.717	0.0483	1.21	0.0010	1.69

The results of the simulations suggested that there could exist such conditions under which the SLI could perform better. Since the standard deviation in the Gaussian model is related to the pixel intensity profile of the point source in the image plane, which in turn has an impact on the accuracy of the SLI algorithm, an appropriate scenario was constructed. Its objective was to determine the SLI's behaviour as a function of the standard deviation σ at constant values of the reference offset δ_{μ}. In this scenario, the $error$, as defined in (8), was computed for each value of the reference offset $\delta_{\mu} \in \{0.1, 0.2, 0.3, 0.5\}$, while sweeping the value of standard deviation over the range $\sigma \in < 0, 3 >$. This scenario resulted in an interesting finding. It showed that the SLI had an exceptionally low error when $\sigma \cong 1.2$. Its error was close to zero around this value of σ, regardless of the value of the reference offset δ_{μ}. The results are shown in Figure 21. It

shows that the error for different values of δ_μ intersects at $\sigma \cong 1.2$. Thus, this point can be considered the optimum operating point for the SLI.

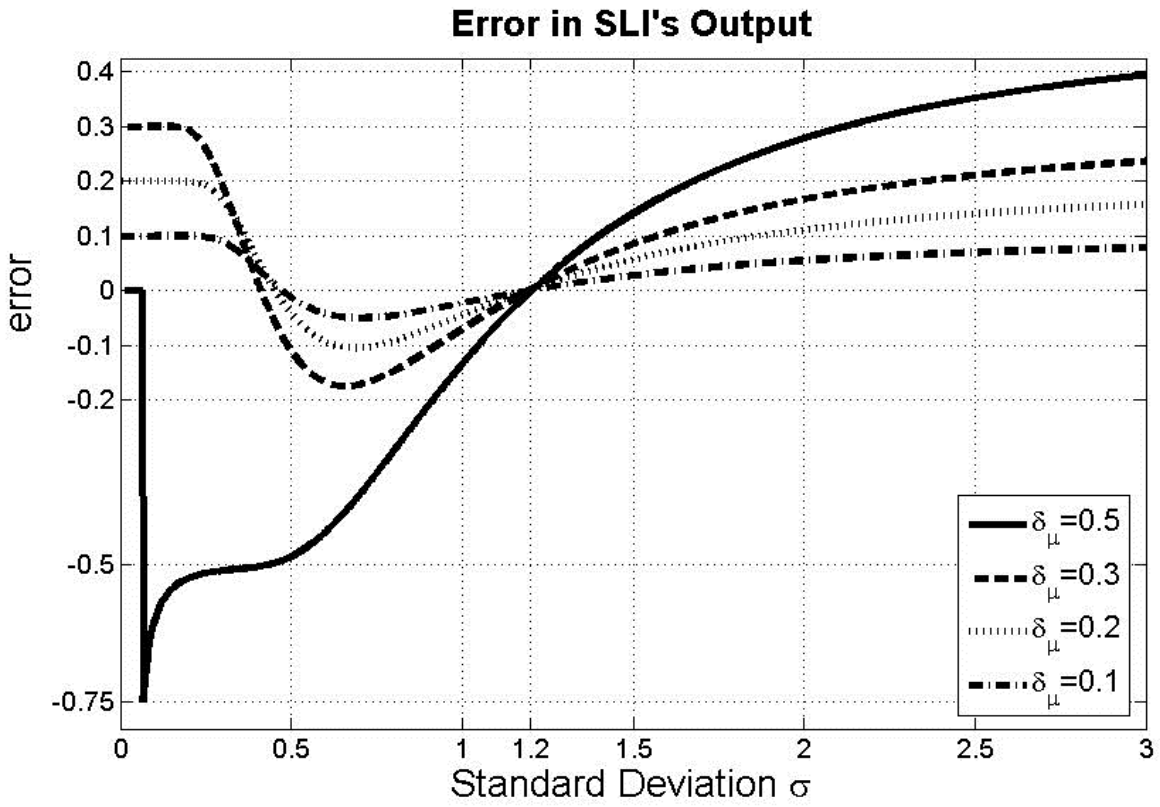


Figure 21: Simulated Error of SLI as a function of σ for different values of δ_μ; error was capped at -0.75

This finding was further supported by repeating the above simulation and computing the output of the SLI, i.e. δ_x for all values of δ_μ, while maintaining the standard deviation constant at $\sigma = 1.2$. A plot of δ_x versus $\delta_\mu$is shown in Figure 22. This plot shows that the $error$ in SLI is very low for all values of δ_μ. Moreover, the computed RMSE in this case was $RMSE = 0.0026$. These results prove that it is indeed the optimum operating point for the SLI. The next section describes the experimental work that was aimed at recreating these conditions for the SLI in the real world and validating the simulated results. The practical implication of this finding is that it is possible to achieve higher accuracy in point detection applications with a low-resolution cameras with the SLI algorithm by ensuring that the pixel intensity profile of the point on the image has the property of a Gaussian distribution with $\sigma \approx 1.2$. These properties can be achieved

through a careful system setup, i.e. an appropriate consideration of elements such as the camera, LEDs and the distances between camera and the LEDs, to name but a few. In practical terms, these findings show that the SLI along with the appropriate system setup can be an enabling factor for many low-power point tracking applications where factors such as physical size, computational power, execution speed, wearable form-factor, are an important consideration. The next section describes the validation process of these findings.

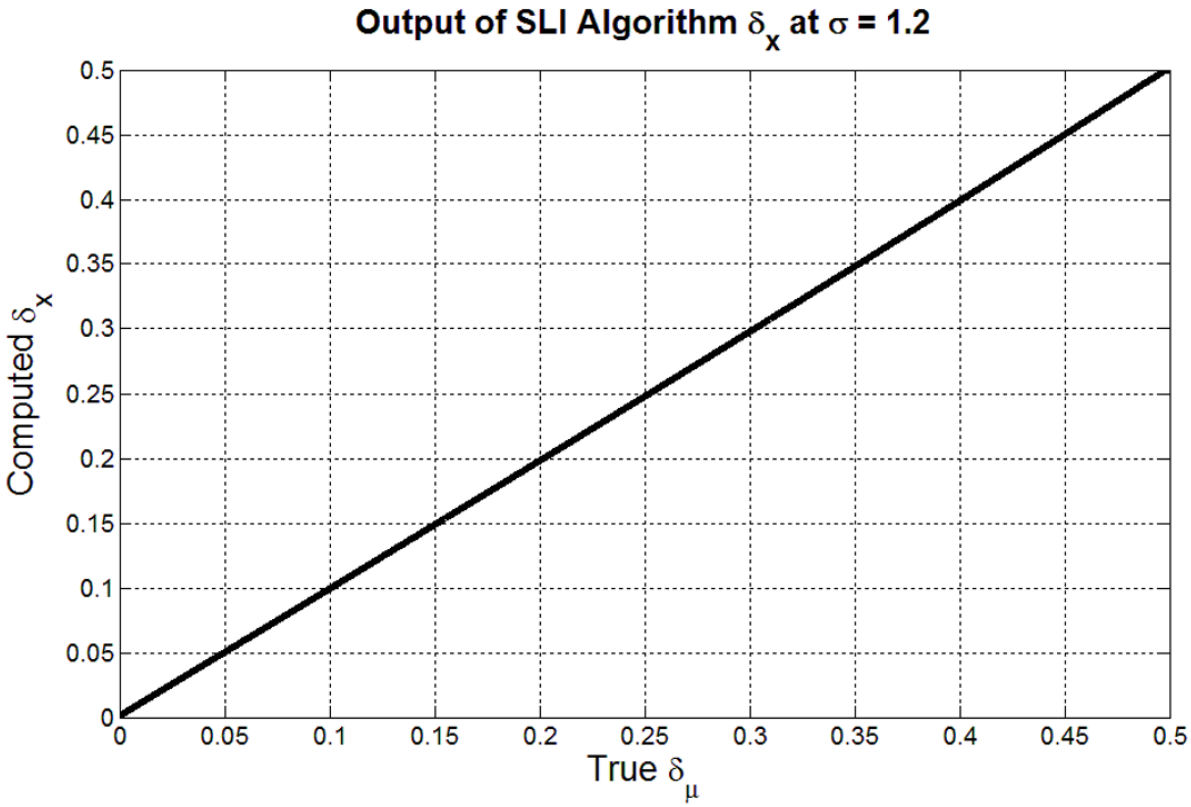


Figure 22: Simulated Output of SLI vs Reference Offset, at constant σ=1.2

3.4 Experimental Validation – Data Acquisition Setup

The experimental setup was designed to create such conditions in the system that the properties of the points of light detected on the image array were as close to the optimal conditions for the SLI method as possible. It was critical to ensure such conditions were created, because the SLI is known to underperform under unfavorable conditions [16].

The system was divided into two major components; the Ambient Environment (AE) and the Wearable Platform (WP), as shown in Figure 23. These elements could be considered as two

nodes in a local network that are coupled together in more than one way. Firstly, the WP was fitted with an IR Long Pass Filter, i.e. passing long IR wavelengths, with high transmittance in the IR wavelength spectrum. The matching IR LED was selected for the AE node. Therefore, the noise floor was lowered and uniform over the imaging device's pixel array. Secondly, the IR LED was carefully selected. Not only did the IR LED have to match the IR Filter, but also its detected point of light on the pixel had to have the required characteristics which was the dynamic intensity range which matched the pixel array's sensitivity. Finally, the two nodes were coupled via a direct wireless telecommunications link. It provided the ability to control the intensity of the IR LED from the WP in real-time. Thus, the intensity of the IR LED could be maintained at the optimal level; regardless of the relative positions of the two nodes. The intensity of the IR LED was controlled to offset the effects of motion of the WP; especially those related to the variation in the distance between the WP and IR LED. To this end, the MCU on the WP could send appropriate commands to the Ambient Environment node, which would decode them and use the Pulse-Width Modulation (PWM) to finely adjust the intensity of the IR LED.

Figure 23: Ambient Environment and Wearable Platform Coupling – Overview

In terms of power consumption, the wearable devices have generally the largest constraints since they must be generally small, light, and, thus have limited computational and power resources. On the other hand, the IR LEDs can be mounted in the ambient environment, e.g. in walls or the ceiling. Therefore, the IR LEDs are less constrained in terms of power consumption; albeit they do not generally require as much power.

3.4.1 *Wearable Platform for 3-D Point Estimation*

The experimental WP was designed and implemented with both the practical application and the experimental requirements in mind. From a practical point of view, the key challenge was to select the most suitable imaging device, or in simpler terms the camera. The most suitable camera for this work was identified and selected. It was the OV8865 made by OmniVision, Inc [92]. It is a state-of-the-art image sensor for low-power, high-performance, mobile applications. Apart from the suitable physical properties, such as energy efficiency, it offers high performance in low-light conditions. A high resolution of this camera, i.e. 3264x2448 was needed for the experimental validation procedure.

The computing platform for this work was the Microsoft Surface Pro 4 tablet computer [93], which housed the OV8865 camera module. This selection was motivated by the practical aspects of the experimental work. It is a fully featured Windows 10 machine with high processing power in a light-weight form factor. It enabled a convenient mounting on a camera tripod. Figure 24 shows the WP with the major elements identified. Furthermore, this tablet computer could support popular scientific or engineering software packages, such as MATLAB. Therefore, it was a suitable choice to implement the functionalities of the WP whilst enabling the tasks related to the experimental work. The Long Pass IR Filter from Edmund Optics, Inc. was selected as the IR Filter [94]. The transmittance of over 90 % in the wavelength range from 800 nm and 1100 nm was a good match for the low power IR LED; described in the next section.

The radio link between the two nodes was implemented using the HM-11 BLE module System on Chip CC2541 from Texas Instruments, Inc. [95]. The commands for the IR LED intensity values were encoded in simple packets and transmitted over a wireless Serial connection directly from MATLAB code to the BLE module.

Figure 24: Wearable Platform

3.4.2 *Ambient Environment for Validation Trials*

The Ambient Environment node consisted of three major parts: an RF module, the MCU, and the IR LED, as shown in Figure 23. The most suitable IR LED for this work was the High Speed Infrared Emitting Diode VSMB11940X01 from Vishay, Inc. [96]. The radiant power peak of this part was centered at 940 nm. The small physical size along with the relatively high radiant power and a wide angle of half-intensity, ± 75 ° ensured a good match with the OV8865 camera module. This IR LED is shown in Figure 25 (a).

The IR LED was directly controlled by an STM-32F401RE based MCU development board and the ARM Mbed development environment [97]. The development board used the same radio module as the WP, i.e. the HM-11 module. The control unit is shown in detail in Figure 25 (b). The intensity control of the IR LED was carried out using a standard 8-bit PWM technique, which provided sufficiently high granularity.

Due to the low power rating of the IR LED, i.e. the forward current, there was no need to design any additional amplifier circuitry at the PWM outputs of the MCU. The STM32F401RE MCU was able to source up to 25 mA per output port [97], which was sufficient to drive the IR LED. Hence, only current limiting resistors were, as shown in Figure 25 (b).

The Ambient Environment node was designed to include additional functionalities for future work. It was designed to support up to four IR LEDs.

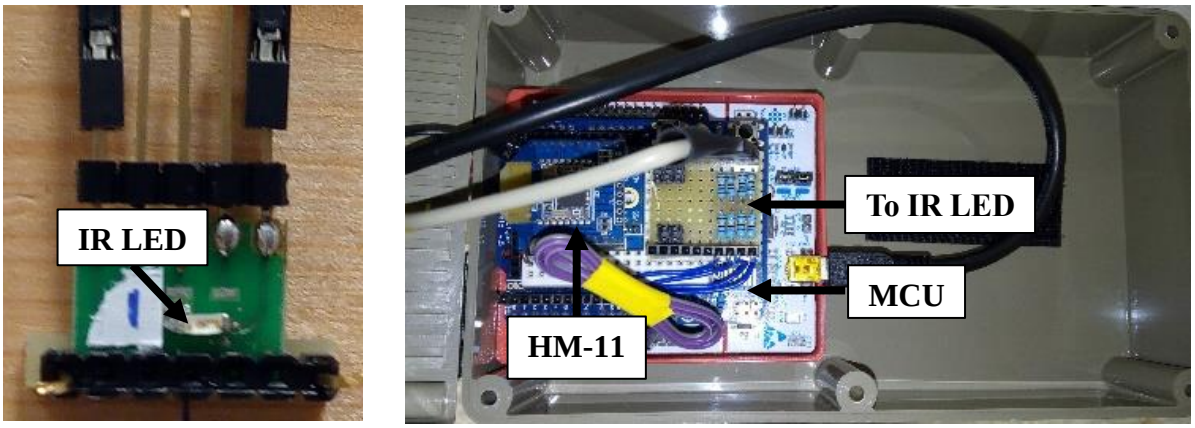


Figure 25: Ambient Environment: (a) IR LED (b) IR LED Control Unit

3.4.3 *Complete Experimental Setup*

The complete experimental setup is shown in Figure 26. It is the equivalent of the general diagram shown in Figure 23. Careful consideration was given to the ambient lighting conditions in the laboratory. Although the IR Filter was used to suppress the undesired ambient light, it could not suppress all sources of light. In particular, the high intensity lamps embedded in the ceiling had the potential to negatively impact the system if they were directly pointed at the camera and were within its FoV. For this reason, these lights were switched off during the experimental work. It resulted in two lamps out of the total of four in the laboratory being off. Although effectively 50 % of the lamps were off, the laboratory remained well illuminated, sufficient for the experiment to take place, as is visible in Figure 26.

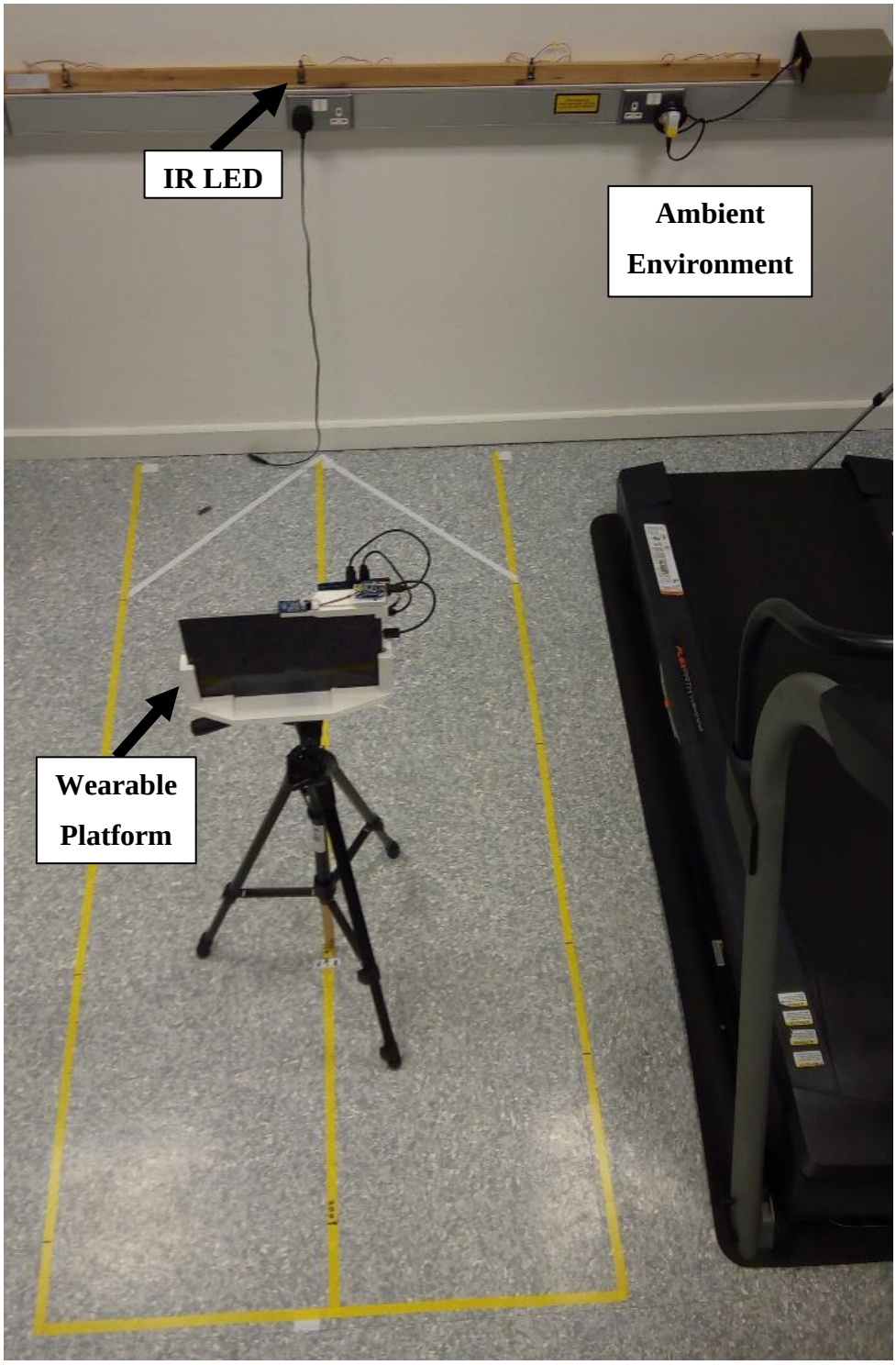


Figure 26: Experimental Platform Implementation

3.4.4 *Work Envelope for Experimental Test Positions*

Work envelope is defined as an area within which the WP can operate and the camera can track the IR LED. The physical work envelope for the experimental work was designed with several factors in mind, as shown in Figure 27. The objective of the work envelope was twofold. Firstly, it had to be large enough for practical applications. Secondly, it had to be such that the optimum conditions for the SLI were maintained in all positions. This was achieved by fully leveraging the capabilities of the camera in the WP and the IR LED. The FoV of the camera approximately overlapped with the radiant intensity characteristics of the IR LED, i.e. the wide angle of half-intensity. Also, the maximum distance between the two nodes was such that the point of light from the IR LED could be detected by the camera; under the optimum conditions for the SLI.

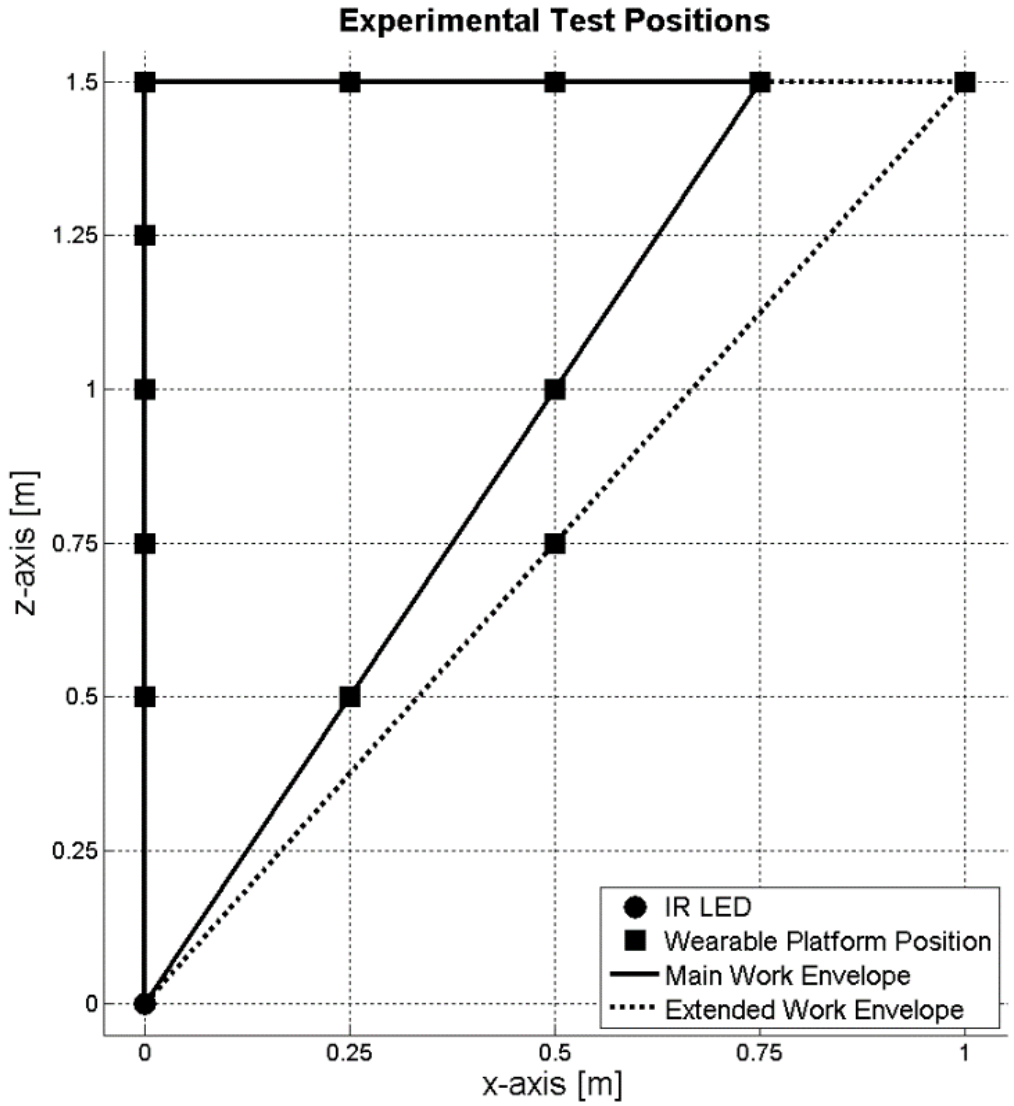


Figure 27: Work Envelope

A series of preliminary tests were carried out to establish the exact limits within which the system could operate in such conditions. The results of these tests were used to establish the optimum work envelope, as shown in Figure 27. This was designed in a grid system that best matched the extreme positions to take measurements at all keep positions. The solid right-angle triangle formed the Main Work Envelope, within which the optimum conditions for the SLI could be maintained. The global frame of reference was a right-handed system with the origin at the IR LED. Because the system was symmetrical about the z-axis, i.e. the Main Envelope could be rotated around the z-axis and the system would maintain its characteristics, this work envelope was sufficient to experimentally validate the SLI method. The position of the WP was defined in three dimensions (3D) as $P = [x, y, z] = [x, 0, z]$. The coordinates x, y and z describe the position of the WP in the World reference frame, which right-handed and its origin coincides with the centre of the IR LED. The y-coordinate was irrelevant in this work, because the WP was always positioned at the same height as the IR LED. Therefore, the y-coordinate was set to zero.

The optimum range of distances along the z-axis was $z \in < 0.5, 1.5 >$ metres. The lateral range of distances along the x-axis at the maximum value of z was $x \in < 0, 0.75 >$, for the Main Work Envelope. An Extended Work Envelope was added by increasing the lateral range to $x \in < 0, 1 > metre$. Thus, the angle between camera's optical axis (or in in other words the z-axis) and the line segment between the IR LED and the camera is increased from 26.57 degrees to 33.69 degrees. It needs to be noted that the camera's optical axis was aligned with the z-axis as closely as possible, while its horizontal axis was aligned with the global x-axis during the experimental data acquisition procedure.

3.4.5 *Data Acquisition Procedure*

The data acquisition procedure in this work was more complex than simulations. Whereas the simulations allowed for the use of artificially created data sets, this experimental work involved the acquisition of data from real input camera frames. The nature of this process had its practical implications, such as the acquisition of large numbers of high-resolution images under controlled conditions.

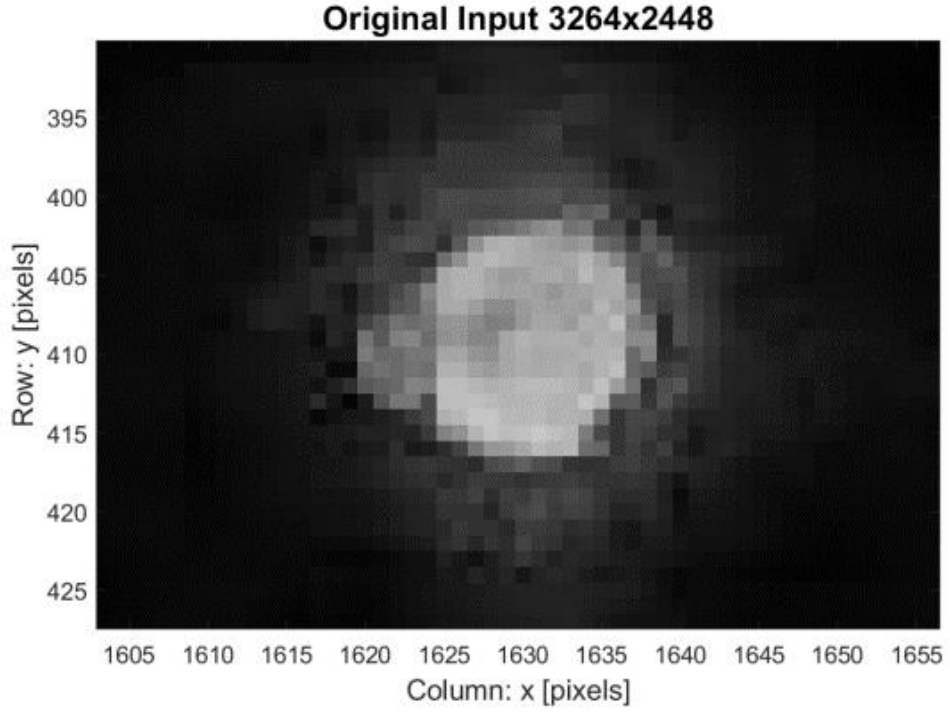
3.4.5.1 Raw Image Frames Acquisition:

The raw input images were acquired with the resolution of 3264x2448 pixels; using the Tablet's built-in Camera application. This application allowed for the full use of the camera module's resolution, as opposed to MATLAB®, which supported only selected resolutions. For each

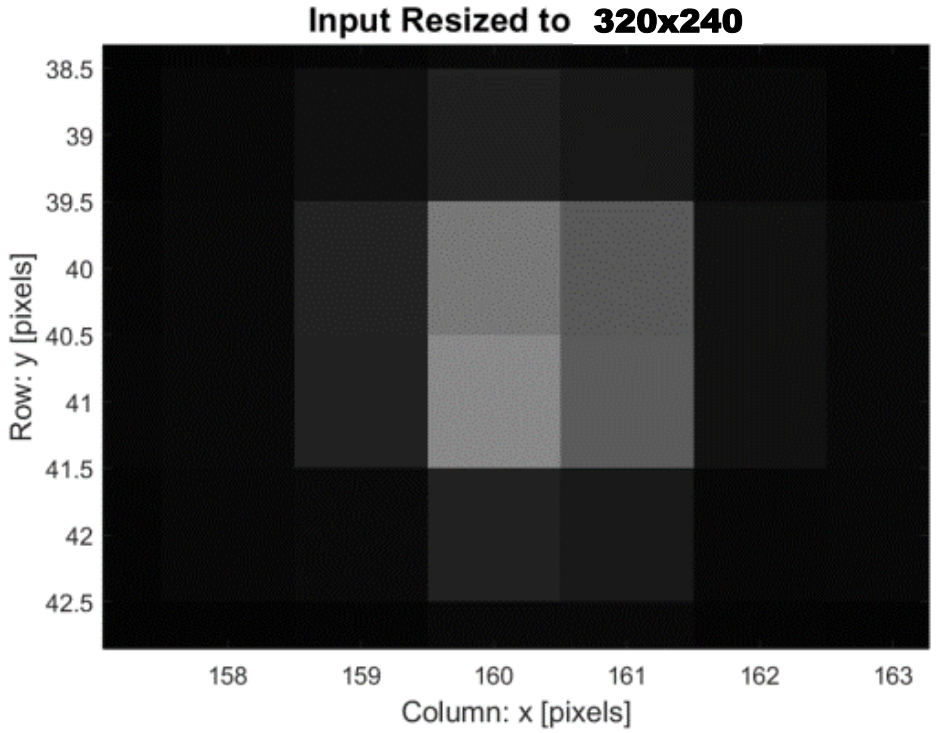
position of the WP in the Work Envelope, N input frames were acquired. Due to the practical implications, in particular the fact that each raw input frame had an average size of 2.5 MB in memory, the number of frames acquired at each position had to be limited. Larger runs of continuous image acquisition caused an instability of the Windows 10 operating system. On the other hand, multiple runs of image acquisitions were not considered, because they could potentially invalidate the results. In order to maximise the chance of achieving constant conditions in the experiments, the most reliable way to validate the SLI was to use a single data set that was acquired in one go, as opposed to multiple acquisitions separated by considerable time intervals. To this end, the resultant number of raw input frames was $N > 1000$, for each position. Once the raw frames were acquired, they were cropped by removing top and bottom 33 % of the images; to avoid having to store and process redundant regions of the images. Since the y-coordinate of the WP's position P was always equal to zero, this operation had no impact on the processing of the point of interest. The point of interest was always located near the vertical centre of the image.

3.4.5.2 *Low Resolution Input Frame Creation*

The second stage of the data acquisition was to use the raw input frames to create the input image for the SLI algorithm. Since the target camera resolution for such application spaces should be as low as possible, while preserving the necessary information about the scene, the resolution was selected to match the standard 320x240 values. The cropped high-resolution input images were resized by down-sampling using the bicubic interpolation algorithm [98, 99]. The input and output of this stage is shown in Figure 28 which is an example of how numerous pixel values from the original high-resolution input image were used to compute the low-resolution output image. It is worth noting that the input images were not resized to exactly 320x240 pixels. The cropping process, where in the top and bottom 33% of the raw images were removed, distorted the aspect ratio of the original resolution. Hence, the input was resized to 320 pixels horizontally, while maintaining the aspect ratio of the cropped input image, as shown in Figure 28 (b).



(a)



(b)

Figure 28: Image Resizing Process - Point Peak in Zoom In: (a) Original Input, (b) Input Resized to Resolution 320x240 using Bicubic Interpolation

3.4.5.3 *Reference Point Estimation*

The peak detection at pixel level was a straightforward task; in the low-resolution images. Under the correct conditions, i.e. using an IR Filter and as single source of light in camera's FoV, the point detection procedure involved finding the maximum pixel value in the image which was the peak of the point of light originating from the IR LED. The pixel intensities at and around these coordinates on the image array were used in the validation work of the SLI algorithm.

The reference point refers to the coordinates of the true peak of the light point that is used to determine the accuracy of the subpixel point detection methods. The reference point was determined using a custom method, which allowed for a direct validation of the SLI algorithm and benchmarking it against the two relevant methods found in the SOA, i.e. the LA and GA. This method is described in detail in the following paragraphs.

The direct validation of a subpixel point detection algorithm can be achieved using a single imaging device. A high-resolution camera can be used in this task. For example, a high-resolution camera can be used to acquire N input frames, which then can be down sampled by a certain factor and used as inputs to the subpixel point detection algorithms. The coordinates of the peaks in the original high-resolution image can be used as the reference points in the validation process (after appropriate scaling) if the ratio between the high-resolution and low-resolution images is high enough, e.g. a factor of ten or more. While this is an effective approach, its reliability or repeatability cannot be trusted. Some empirical testing can show that the pixel intensities at the peak's location can vary greatly from frame to frame. Thus, the location of the peak may shift from frame to frame. It is unacceptable, because the location of the reference point should not vary during the experiment if both camera and the point source of light are stationary. This issue is particularly apparent in low-light conditions with IR LED and cameras fitted with a matching IR filter. Our experimental setup consisted of an IR LED and an 8 Megapixel camera with a matching IR filter, both of which were stationary in controlled laboratory conditions. The fluctuations in pixel intensity were observable with naked eye when zooming in onto the point. This problem shows the need for a more reliable way of determining the point of reference.

The problem of fluctuations in pixel intensities in the successive frames could be tackled with statistical methods, e.g. by computing mean peak location over all N input frames. One way of finding the peak's location can involve using the Circular Hough Transform (CHT), which can find the centres of a circle in images [100]. The circle centres should be coincident with the location of the points' centres, assuming a symmetric Gaussian distribution of the pixel intensities

around the peak. The mean of the circle centres from the entire set of N input images can be computed and used as the reference point. However, it may prove to be insufficiently accurate. Although the intensity peak does generally have the form of a circle, it is not always an ideal circle. The shape of the pixel distribution can vary due to noise and angle at which the light rays intersect the pixel array. Also, the pixel intensity profile in low-light conditions does not necessarily have pure Gaussian properties, as shown in Figure 28 (a). That is, the circle centre does not always coincide with the intensity peak. An alternative approach can involve calculating the mean of peak intensity locations over all N input frames. Again, it may not be a reliable measure as the standard deviation of this metric would be high, given the fact that the high-intensity pixels can be spread over a relatively large area.

The proposed algorithm was designed as a multi-step iterative process to determine the reference point. The location of the peak was estimated based on a combination of mean CHT and mean peak pixel intensity over all N input frames. The estimate of the true peak's location $\tilde{P}$ was updated in each iteration of the loop. The algorithm was run over n iterations until the $\tilde{P}$ no longer changed, i.e. the best achievable solution for this method was determined. The algorithm is shown in **Error! Reference source not found.**. All variables used in this algorithm are two-e lement row vectors with the elements corresponding to the x- and y-axis, as shown below in (11):

$$\tilde{P} = [x\ y] \tag{11}$$

The first step of the reference point estimation algorithm involved the acquisition of a relatively large set of N input frames from the high-resolution camera in an experimental setup. It is assumed that the experimental setup is placed in a controlled laboratory environment. It is critical to ensure that there are no external light intensity fluctuations originating from uncontrolled ambient light which can be a source of noise. Secondly, there must not be any mechanical vibration present in the environment during the data acquisition process. Any mechanical distortion that could cause motion of either the camera or the source of light should be avoided, e.g. a slamming door, air drafts caused by motion or air conditioning systems, or even loud talking near the camera.

The second step involved finding an initial estimate of the peak's location $\tilde{P}$. It served as the initial input to the algorithm's loop to be refined over n iterations. It was determined by finding the location of the peak intensity.

In the third step, the Region of Interest (ROI) was set. The centre of the ROI was set to the current value of $\tilde{P}_n$. The size of the ROI could be set manually. It depended on the resolution of the camera and the size of the point on the pixel array. The size of the ROI had one primary requirement. Its area had to be greater than the area of the peak on the pixel array. The specific settings depended on camera's resolution and the size of the point on the pixel array. In this work, the size of the ROI was set to 50 pixels along x-axis and 35 pixels along the y-axis, as shown in Figure 28 (a).

The next three steps were aimed at finding the best location estimate for the given iteration of the loop, n. Steps four and five were focused on finding the mean peak locations using two different methods. Firstly, the mean circle centre $\bar{C}$ was computed from all N input frames in the ROI. The result should be coincident with the peak's location since its 2D pixel intensity distribution forms an approximate circle around it. The CHT algorithm proposed by Atherton et al. was used in this step [100]. Subsequently, the mean pixel intensity $\bar{P}_{intens}$ was computed within the ROI. Finally, the peak estimate, for the given loop iteration n, $\bar{P}$ is the mean of $\bar{C}$ and $\bar{P}_{intens}$. It is also the mid-point between these two values. Optionally, a weighted mean could be considered if either of the two parameters was considered more important, or accurate, in the calculations.

The final stage of the algorithm was aimed at determining whether the mean estimate $\bar{P}$ was less than half a pixel away from the current estimate $\tilde{P}_n$. If so, the algorithm's work was complete. It could proceed to the next steps, i.e.: setting the peak's final position estimate $\tilde{P}$, appropriately downscaling it, and down sampling all N frames in the input set. Otherwise, the process must: go back to Step 3, adjust the ROI_{centre} with new, more accurate position estimate $\tilde{P}_n$, and execute the next loop, $n + 1$. The execution continued until the condition in Step 7 was false and the best estimate of the centre was identified.

TABLE 4: REFERENCE POINT ESTIMATOR ALGORITHM

Algorithm 1: Reference Point Estimator

1. Acquire N high-resolution input frames
2. Get initial peak estimate $\tilde{P}_n$; $n = 0$
3. Set $ROI_{centre} = \tilde{P}_n$
4. For all N frames, find $\bar{C}$
5. For all N frames, find $\bar{P}_{intens}$
6. Find mean estimate $\bar{P} = Mean(\bar{C}, \bar{P}_{intens})$
7. $If\ Abs(\tilde{P}_n - \bar{P}) > 0.5\ pixel$
    a. $n = n + 1$
    b. $\tilde{P}_n = \bar{P}$
    c. $GOTO$ Step 3
8. $\tilde{P}_n = \bar{P}$
9. Down-scale $\tilde{P}$
10. Down-sample all N input frames

This method was validated in experimental conditions with the same setup as that used in validating the SLI algorithm; described in this chapter. A more detailed description of this algorithm and its validation can be found in our published work on this topic [101].

3.5 Results and Discussion

The SLI was experimentally evaluated using the exact same methodology as that used in the development of the simulated modelling. Once the reference point for a given test position was determined, and the pixel intensities of the peak and its surroundings were extracted, the validation procedure from the simulation could be readily applied. The experimental work was divided into two scenarios:

- Scenario 1: Intensity of the IR LED was constant for all test positions.
- Scenario 2: Intensity of the IR LED was dynamically controlled to maintain optimum conditions for the SLI algorithm

Apart from the above differences, the overall procedure was the same for both scenarios. The SLI was validated using datasets acquired at each test position in the Work Envelope.

The results were compiled in tables compatible with those in the simulations to allow for a side-by-side comparison. The rows were extended to contain the results obtained from data acquired at different test positions. The results were grouped and tabulated to show the performance along the main axes in the Work Envelope. This way, the changes in performance of the system could be clearly observed, as the WP was moved to different test positions.

3.5.1 *Scenario 1– No LED Intensity Control*

The results of this scenario, for the test positions: along z-axis, x-axis, and at the remaining diagonal positions, are shown in: Table 5, Table 6, and Table 7, respectively.

The first observation, that one can make, in terms of the execution time, is that results unveil a pattern similar to that of the simulations, as regards to the execution time. The SLI required the least amount of time to find the peak at sub-pixel level, which was expected. Its average execution time was $2.4\ \mu s$. The LI was slower, as it required an average of $3.07\ \mu s$ to complete this operation, which was also expected. However, the execution time of the GA method was unexpected. It turned out to be slightly faster than the LI, with the average time of $3.01\ \mu s$.

The overall results in terms of the accuracy did not conform to literature expectations. The SLI was expected to underperform, compared to the remaining two methods. Given these conditions, the SLI performed better than other two methods. The Root Mean Square Error (RMSE) values of the remaining two methods were high. This scenario was designed specifically to reveal and explore the shortcomings of the SLI method; to try to identify the best possible conditions for it. As a by-product, it was found that both the LI and GA also underperformed under these conditions. There results revealed that the LI and GA were not as robust. However, the investigation as to why this was the case was secondary and beyond the scope of this work.

The RMSE values for the SLI, shown in Table 5, reveal an interesting, though expected, behavior. The RMSE of the SLI decreased considerably with the distance between the WP and the IR LED. The impact of the constant intensity of the IR LED, as the distance varied, was clearly visible, since these results were obtained at test positions $P = [0,0,z]$ where $z \in < 0.5,1.5 >$. These results were in line with expectations, as the peak pixel intensity detected on the imaging sensor was expected to vary with the distance z, which was the only variable parameter in this case. The imaging sensor was initially saturated at the lowest value of z, hence the very high values of the RMSE. Subsequently, the RMSE values decreased, as the value of z increased, thus decreasing the perceived peak intensity on the imaging sensor's array. These findings demonstrated that, the

intensity of the IR LED had a significant impact on SLI's performance, in this system. Furthermore, these results suggest that the intensity of the IR LED could be used to further optimize the conditions for the SLI.

The results obtained from test positions along x-axis and those at the diagonal positions are shown in Table 6 and Table 7, respectively. These results further support the findings described in the previous paragraph. Apart from that, they also demonstrate the extent of the impact of the angular displacement on the SLI. The angular displacement, i.e. the angle between the optical axis of the camera and the line segment between the camera's principal point and the IR LED, has a significant impact on the performance of the system. These results show that SLI had lower accuracy at lateral test positions which was particularly evident at the extreme test positions; in the Extended Work Envelope.

TABLE 5: RESULTS: SCENARIO 1; TEST POSITIONS ALONG Z-AXIS

	SLI		**LI**		**GA**	
	RMSE $[\boldsymbol{pixel}]$	***Time*** $[\mu s]$	$\boldsymbol{RMSE[pixel]}$	***Time*** $[\mu s]$	$\boldsymbol{RMSE[pixel]}$	***Time*** $[\mu s]$
0,0,0.50	**0.5000**	2.4383	0.4795	3.6264	0.4879	3.0681
0,0,0.75	**0.4830**	2.5074	0.3357	3.1586	0.3326	3.0036
0,0,1.00	**0.3039**	2.6114	0.3925	3.6640	0.3868	3.0885
0,0,1.25	**0.2750**	2.4075	0.3640	2.9063	0.3587	2.8925
0,0,1.50	**0.2744**	2.2868	0.4080	2.8603	0.3946	3.1092

TABLE 6: RESULTS: SCENARIO 1; TEST POSITIONS ALONG X-AXIS

	SLI		**LI**		**GA**	
	$\boldsymbol{RMSE[pixel]}$	***Time*** $[\mu s]$	$\boldsymbol{RMSE[pixel]}$	***Time*** $[\mu s]$	$\boldsymbol{RMSE[pixel]}$	***Time*** $[\mu s]$
0.00,0,1.5	**0.2744**	2.2868	0.4080	2.8603	0.3946	3.1092
0.25,0,1.5	**0.2028**	2.3393	0.3195	2.7439	0.3049	2.7724
0.50,0,1.5	**0.3327**	2.2432	0.3677	2.8863	0.3640	2.8321
0.75,0,1.5	**0.2481**	2.4142	0.3600	3.2558	0.3710	2.9696
1.00,0,1.5	**0.3706**	2.5887	0.3957	2.8967	0.4226	2.9662

TABLE 7: RESULTS: SCENARIO 1; DIAGONAL TEST POSITIONS

	SLI		LI		GA	
	$RMSE[pixel]$	*Time* [μs]	$RMSE[pixel]$	*Time* [μs]	$RMS[pixel]$	*Time* [μs]
0.25,0,0.5	**0.1472**	2.3239	0.2126	2.8268	0.2350	2.8105
0.5,0,1.00	**0.2554**	2.4517	0.2610	3.2611	0.2716	2.8442
0.5,0,0.75	**0.3711**	2.3203	0.2886	3.0136	0.2659	3.6208

3.5.2 *Scenario 2 - LED Control for Optimum Conditions*

The objective of this scenario was to evaluate the performance of the SLI under the optimum conditions. The optimum con ditions for the SLI are such that the ratio of pixel intensities at the detected point's peak is as fallows $\frac{f(x)}{f(x-1)} \cong \frac{1}{2}$, as shown in Figure 20 in section 3.2.2. To this end, the intensity of the IR LED was controlled at each test position in the Work Envelope. The results from Scenario 1 were analyzed to determine the conditions, under which the SLI performed best. The raw input images acquired at the test positions where the SLI had the lowest RMSE, were analyzed in detail. The pixel intensity profiles the peaks in these input images revealed that the maximum intensity values of the peak $I(x, y)$ tended to have similar values. Furthermore, the intensity value distribution at and around the peak tended to have Gaussian properties; that were close to those identified as optimum for the SLI. An additional analysis was carried out to determine what peak pixel intensity values $I(x, y)$ yielded pixel intensity profiles with the optimum Gaussian distributions. The most suitable peak pixel intensity value in the input image was $I(x, y) \cong 70$ (in the 8-bit range). However, for practical reasons, a range of values was selected, $I(x, y) \in < 63{,}76 >$. Therefore, the data acquisition procedure was preceded by one additional step. For each test position, the intensity of the IR LED was set with such a PWM value that the measured peak pixel intensity in the resultant input images were within the identified range of values. Once that PWM was determined, the experimental procedure continued.

As in Scenario 1, which is described in the previous section, the results from all test positions were grouped in three tables: along the z-axis, x-axis, and at the diagonal positions, as shown in: Table 8, Table 9, Table 10, respectively.

The overall execution times of the three methods were comparable to that in the Scenario 1. There were subtle differences. The average execution time of the GA increased to $3.25\ \mu s$, whereas that of the LI decreased to $2.98\ \mu s$. The SLI had the shortest average execution time with $2.18\ \mu s$.

The optimization of the intensity of the IR LED had a significant impact on the accuracy of the three methods. The RMSE values significantly decreased across both all methods and test positions. The results of the LI and GA were much closer to the expected values. Though, the RMSE values for the GA show that it is not as robust in this scenario either. These results were probably affected by noise in the system, wherein the assumed ideal Gaussian distribution in the pixel profiles was more difficult to achieve.

TABLE 8: RESULTS: SCENARIO 2; TEST POSITIONS ALONG Z-AXIS

	SLI		**LI**		**GA**	
	***RMSE*[*pixel*]**	***Time*** [μs]	***RMSE*[*pixel*]**	***Time*** [μs]	***RMSE*[*pixel*]**	***Time*** [μs]
0,0,0.50	**0.0703**	2.1197	0.1025	2.8236	0.1219	3.5245
0,0,0.75	**0.1138**	2.5858	0.1577	2.7908	0.1627	3.7434
0,0,1.00	**0.1353**	2.1651	0.0390	3.0422	0.0415	2.9283
0,0,1.25	**0.1145**	2.6931	0.0448	3.1981	0.0410	4.0859
0,0,1.50	**0.1903**	1.8682	0.1975	2.8421	0.1981	3.2630

TABLE 9: RESULTS: SCENARIO 2; TEST POSITIONS ALONG X-AXIS

	SLI		**LI**		**GA**	
	***RMSE*[*pixel*]**	***Time*** [μs]	***RMSE*[*pixel*]**	***Time*** [μs]	***RMSE*[*pixel*]**	***Time*** [μs]
0.00,0,1.5	**0.1903**	1.8682	0.1975	2.8421	0.1981	3.2630
0.25,0,1.5	**0.0653**	2.3132	0.1140	3.0252	0.1034	3.5351
0.50,0,1.5	**0.1546**	1.9154	0.1504	3.0281	0.1497	3.0463
0.75,0,1.5	**0.1510**	1.9141	0.1525	2.8461	0.1551	2.9978
1.00,0,1.5	**0.3623**	1.9616	0.3164	3.0223	0.2886	2.9526

TABLE 10: RESULTS: SCENARIO 2; DIAGONAL TEST POSITIONS

	SLI		LI		GA	
	***RMSE*[$pixel$]**	***Time*** [μs]	***RMSE*[$pixel$]**	***Time*** [μs]	***RMSE*[$pixel$]**	***Time*** [μs]
0.25,0,0.5	**0.0589**	1.9660	0.0761	3.0034	0.0487	3.0739
0.5,0,1.00	**0.0632**	2.5184	0.7840	2.9866	0.0949	2.8508
0.5,0,0.75	**0.2092**	2.4515	0.2092	3.3436	0.1783	2.9680

Perhaps, the most interesting observation was made with respect to the accuracy of the SLI method. The RMSE values of the SLI were significantly lower, as compared to the corresponding results obtained in Scenario 1. The RMSE values in the Main Work Envelope were within $0.2\ pixel$. Moreover, the SLI achieved the highest average accuracy in the Main Work Envelope, as compared to the other two methods. The average RMSE was $0.11\ pixel$, whereas the average RMSE of the LI and GA was $0.23\ pixel$ and $0.12\ pixel$, respectively. If the results from the Extended Work Envelope are included, the SLI and GA had the same average RMSE of $0.14\ pixel$, while that of the LI was equal to $0.2\ pixel$.

It is worth noting that all three methods achieved lower accuracy, i.e. higher RMSE, at test positions in the Extended Work Envelope. The results from the Extended Work Envelope were expected to be worse in all three methods due to the high angular displacement and its implications on the system. The IR LED becomes increasingly smeared on the pixel array as the WP moves laterally, which distorts the symmetry of the pixel profile, or pixel intensity distribution, of the captured point of interest.

3.5.3 *SLI – Performance Analysis*

A closer look at the performance of the SLI, in both scenarios, can illustrate the impact of the dynamic intensity control of the IR LED. Figure 29 shows a graphical comparison of the RMSE of the proposed method along with that of the LI and GA along the z-axis. The navy-blue bars (leftmost bar at each test position) represent the RMSE in Scenario 1. It clearly demonstrates the relationship between the RMSE and the distance z, at $PWM = 100\ \%$. Initially the imaging sensor was saturated, thus the maximum error $RMSE = 0.5\ pixel$. As the distance z increased, the RMSE decreased. On the other hand, the results from Scenario 2 show the difference that the optimization of the IR LED's intensity made. The RMSE was significantly reduced, and, more

importantly, maintained at a steady level across the range of values of z. The RMSE increased at $z = 1.5\ m$ to just under $0.2\ pixel$, because, at that distance, it was increasingly more difficult to maintain the peak pixel intensity in the centre of the optimum intensity interval $I(x, y) \in < 63{,}76 >$. The analysis of the RMSE along the x-axis, at $z = 1.5\ m$, further supports this observation, as shown in Figure 30. Although the RMSE was maintained below $0.2\ pixel$ within the Main Work Envelope, the gap between results from the two scenarios is clearly narrower and decreasing as the values of x increased. Finally, the two values almost converged at the edge of the Extended Work Envelope, at $x = 1\ m$. The reason for this convergence is twofold. On the one hand, the radius, i.e. $r = \sqrt{x^2 + z^2}$, was so large, that the PWM control had reached its maximum limit, and was no longer able to maintain the $I(x, y)$ values within the optimum intensity interval. On the other hand, the distortions related to the angular displacement became considerable. At this angle, the pixel peak became smeared and asymmetrical on the imaging sensor; along camera's x-axis. These two factors in aggregate led to higher RMSE of SLI in Scenario 2.

Furthermore Figure 29 and Figure 30 offer the graphical means for comparing the SLI to LI and GA, with respect to the RMSE metric. The first observation is that the dynamic control of the IR LED's intensity had a positive impact on the accuracy of all three methods, as all of them achieved lower RMSE, as compared to their corresponding results in Scenario 1. Moreover, it can be observed that the SLI had the lowest RMSE in Scenario 2 at most test positions.

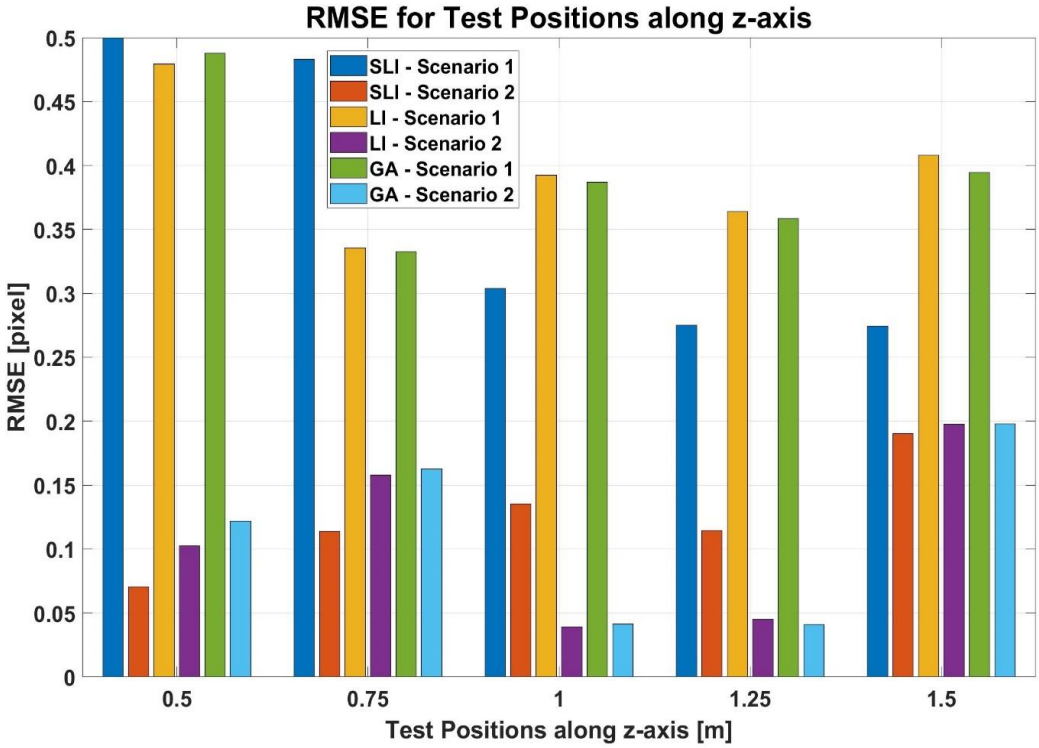


Figure 29: RMSE of SLI, LI and GA along z-axis, at x = 0 m

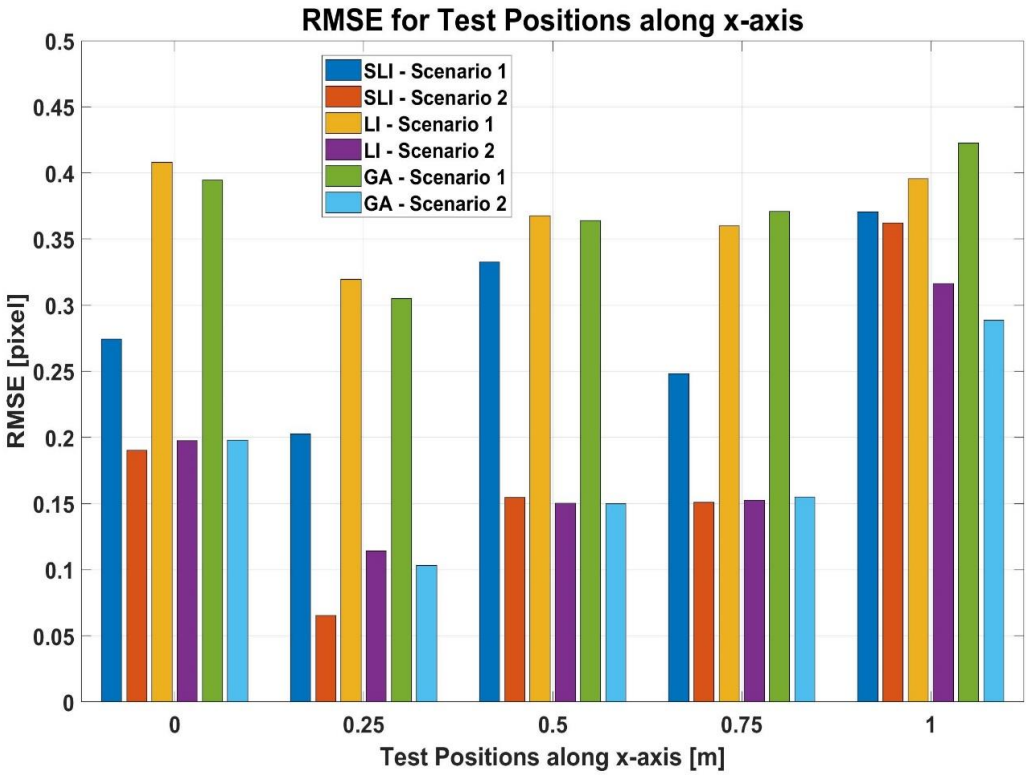


Figure 30: RMSE of SLI, LI and GA along x-axis at z = 1.5 m

A deeper analysis of the pixel intensity profiles of the input images in Scenario 2 can explain the results more comprehensively. The pixel intensities at the peak's location $f(X)$, and the adjacent pixels, used as the input to the SLI, can be used to determine the overall characteristics of the peak. Given the assumed symmetric Gaussian distribution of the intensity profile of the peak, a curve fitting technique can be used to fit a Gaussian model, to inspect its properties. The standard deviation term σ in this model could be compared to the optimum value identified in the simulations, i.e. $\sigma \cong 1.2$. To test that, MATLAB's built-in curve fitting tool was used to fit an n-th order Gaussian model to the intensity profiles, as defined in (12):

$$f(x) = \sum_{i=1}^{n} a_i e^{\left[-\left(\frac{x-\mu_i}{c_i}\right)^2\right]} \tag{12}$$

Given the direct one-to-one correspondence to the standard Gaussian probability distribution model [102], the c_i term in (12) is related to the standard deviation, as defined in (13).

$$\sigma_i = \frac{c_i}{\sqrt{2}} \tag{13}$$

In order to obtain the most comprehensive results, the pixel profiles from the test positions along the z-axis, i.e. $P = [0,0,z]$; $z \in < 0.5,1.5 >$, were used in the analysis. Firstly, at each test position P, the mean pixel intensities were computed, for each pixel coordinate at the peak; using all input images. Then, these mean values were normalized, to ensure that the result remained within the $< 0,1 >$ interval. Finally, the Gaussian model was fitted to the pixel intensities, for each test position P. The most reliable results were obtained with the 1-order model, i.e. $n = 1$ in (11). Although higher-order fitting runs resulted in more precise results, not all peaks were modeled as the single-peak Gaussian distribution. Hence, the results from the 1-order model were used in the analysis. The results are depicted in Figure 31. At first glance, the resulting fits suggest that all peaks from the selected test positions were very similar. Although the range of values of z in the Work Envelope was considerable, the dynamic control of the intensity of the IR LED

proved to be effective at maintaining the optimum conditions for the SLI. The values of the standard deviation σ of each model fit were the key parameters in this analysis. The range of the standard deviation of the fits, as computed using (11), was $\sigma \in < 0.58, 0.93 >$. The mean standard deviation of all fits was $\overline{\sigma} = 0.713$. The comparison of these results with the simulations can show whether the assumptions made at formulating the proposed method were correct and achievable in a practical experimental setup. The simulated values of RMSE for the experimentally measured interval of σ, and the mean $\overline{\sigma}$, can be compared with the corresponding experimental results. According to the simulations, the SLI was expected to have $RMSE \cong 0$ at the standard deviation $\sigma \cong 1.2$. Moreover, the RMSE was expected to remain low around this value of σ. In the case of the measured interval of σ, the maximum error was expected to be $RMSE \leq 0.40\ pixel$, for $\sigma = 0.58$, (Figure 21). The experimentally measured maximum error in the Main Envelope was $RMSE = 0.1903\ pixel$. Furthermore, the error at the $\overline{\sigma}$ was expected to be RMSE $\leq 0.20\ pixel$. The corresponding experimentally measured average error in the Main Envelope was $RMSE = 0.11\ pixel$. These results prove that the experimental results were in line with the predicted performance of the system.

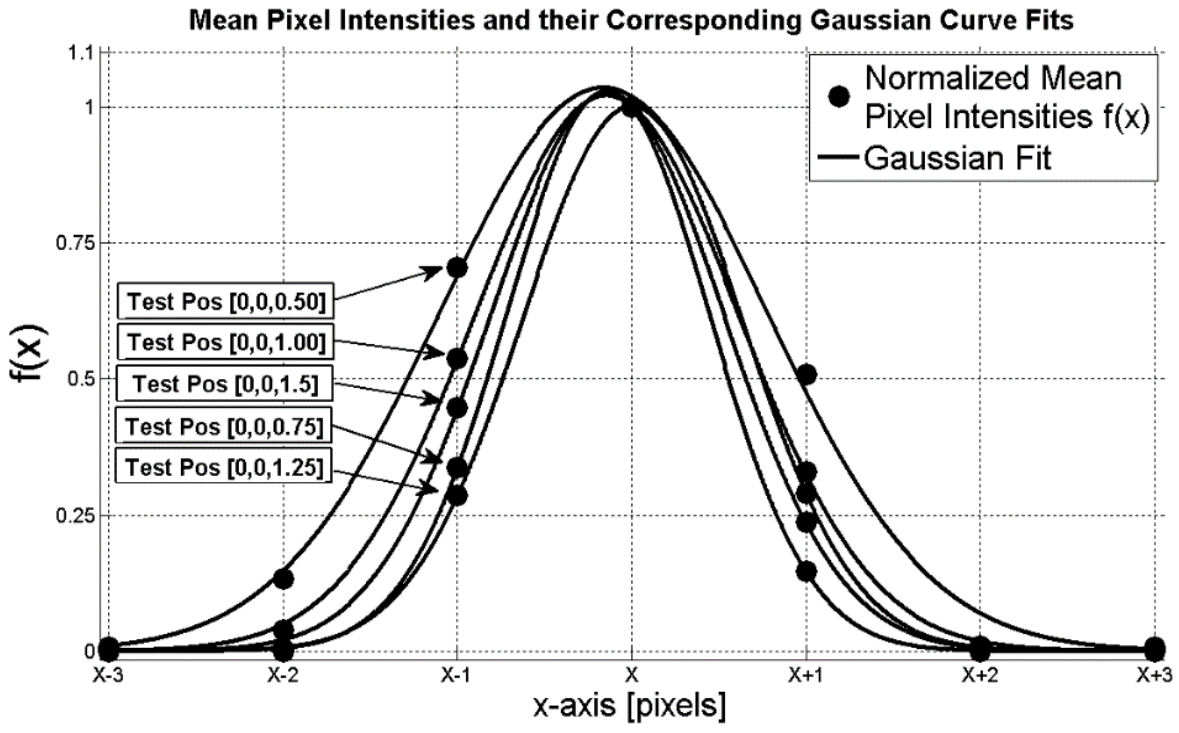


Figure 31: Pixel Intensity Profile Analysis

It needs to be noted that this experimental setup did not produce the exact ideal Gaussian pixel intensity profiles; with the optimum $\sigma \cong 1.2$. Although, the mean $\overline{\sigma}$ was very close to the optimum value, the best theoretically predicted performance of SLI was not achieved. The collective impact of all the contributing factors resulted in difficulties in tuning the system to the exact ideal operating point, i.e. the $\sigma \cong 1.2$ across the entire work envelope. That is not to say it is impossible or impractical. Given these results, it should be possible to further optimize the system to set its operating point closer to the optimum one. The various parameters in the system can be optimized to shift the $\overline{\sigma}$ closer to the optimum value. For example, an IR LED with slightly larger active area, or the reflective element, could help achieve that. Apart from shifting the $\overline{\sigma}$, the spread of the σ values could be also narrowed by a more precise PWM control, for example.

Nevertheless, the obtained experimental results were as expected. The assumptions which were made when developing the proposed algorithm, were proven correct. These assumptions were that the system could be set up such that the peak of the point of interest was distributed across an area of 3x3 to 5x5 pixels and the ratio between the ratio of the pixel location with peak intensity to the preceeding pixel location, i.e. $\frac{f(x)}{f(x-1)} \cong \frac{1}{2}$, could be achieved and maintained approximately constant. Moreover, these conditions should be readily reproducible in a practical application. From a certain point of view, these results are very promising. They prove that the predicted high accuracy, and low execution time, are indeed achievable in a practical application. If the accuracy of the SLI was further increased, it should be possible to decrease the resolution of the camera. It should be possible to at least half the resolution, from $320x240$ pixels, down to $160x120$, while maintaining the accuracy of the point detection. It can have very significant implications on the practical requirements of the WP. The resolution of the imaging sensor is the single parameter of the WP that drives the requirements for the other components in the system. The resolution is directly linked to the amount of the required computations per frame, as well as the frame rate itself. The type of the processor is largely dependent of the computational requirements of the camera and the algorithms that are executed on the images. The battery life, and even the physical size, heavily depend on the computational requirements of the system. In summary, if this single parameter could be decreased by a factor of two, for example, the WP could be physically smaller, lighter, and have a longer battery life. Thus, the proposed system would be even more viable in practical resource-constrained applications.

3.6 Conclusions

The design, implementation, and results of the experimental validation of the SLI algorithm, as well as the formulation and modelling, have been presented in this chapter. This method and its performance in the presented practical experimental setup represent a significant improvement over the current SOA. The SLI outperformed the other two methods, with respect to both accuracy and the execution time; in the Main Work Envelope. The SLI's average RMSE was $0.11\ pixel$, versus $0.23\ pixel$ and $0.12\ pixel$ for LI and GA, respectively. Even if the Extended Work Envelope was considered, the SLI still matched the most accurate method in the SOA, the GA with $RMSE = 0.14\ pixel$. The SLI had the shortest execution time, when compared to the other two methods. The average execution time of the less accurate LI was 37 % higher, and that of the matching GA was almost 50 % higher, than that of the SLI. These results prove that, not only is the SLI as accurate as the best comparable method in the SOA while being much faster, but also that the ideal operating conditions for the SLI are achievable in the context of point detection and tracking with low-power wearable camera systems. The results also show that there is room for improvement, that can further increase SLI's accuracy, at little to no additional expense. Thus, the SLI can be considered as a practical choice for application spaces where, apart from accuracy, other factors such as cost, physical size, or battery life are important.

4 Multimodal Sensor Fusion; Monocular 3D Pose Estimation

M. P. Wilk, M. Walsh, and B. O'Flynn, "Multimodal Sensor Fusion for Low-Power Miniaturised Wearable Human Motion Tracking Systems in Sports Applications," IEEE Sensors, 2020, (under review)

4.1 Introduction

As described in previous chapters, and in particular in section 2.1.2, multimodal sensor data fusion is a common approach to solving problems in applications wherein a single sensor modality fails to provide enough information to solve the given problem. In such cases, sensors with different modalities are often used together to overcome this difficulty. The complementary nature of certain sensor modalities can be helpful for tackling problems that would be difficult to solve otherwise. One of the most common examples include the IMU sensor, which is often considered as a single device whose three sensor modalities are fused together to produce a reliable orientation measurement using an algorithm, such as that based on the Gradient Descent [46]. Another example of such a complementary pair of sensor modalities is the combination of vision sensors and IMU sensors. Vision sensor technology can provide information that the IMU cannot capture and vice versa. For example, the camera can determine the absolute position of a given point. This is difficult to achieve that when using IMU due to its inherent limitations, such as the drift or the disturbances in magnetic field [103]. Likewise, the IMU can capture motion parameters independently of the lighting conditions or occlusions, which are some of the main weaknesses of the vision sensors. Therefore, multimodal data fusion techniques are often considered in motion tracking applications, especially those relying on highly miniaturized, low-power, wearable devices. The fusion of vision and IMU sensor modalities is one of the most common approaches in this context thanks to their complementary nature. The combination of these two sensor modalities, in conjunction with sensor fusion algorithms, can result in a reliable 6-DOF pose detection. One of the most notable advances in the SOA is the inclusion of a camera in the wearable device itself. An example of both sensor modalities embedded in the wearable motion tracking devices for 6-DOF pose detection was proposed by Foxlin et al. [73, 75, 104], including their latest product IS-1500 [11]. These are inside-out tracking systems that use a monocular camera (single camera) to track multiple fiducial markers embedded in the ambient

environment and an IMU to correct for the motion and occlusions [105]. Other examples include outside-in tracking systems where a monocular camera was embedded in the ambient environment to track two points of reference attached to a mobile/moving device; that also incorporated an IMU [12, 13].

These works show the evidence for an emerging trend in 3-D pose detection methods that increasingly incorporate monocular vision and IMU sensors in a single wearable unit. The wearable unit is effectively a wearable smart sensor that is driven by the multimodal sensor fusion algorithms. The advances in the SOA in camera miniaturization [14, 78], are accompanied by algorithms that can detect the precise location of points of interest at subpixel level, thus allowing for a lower resolution of the camera [91, 106], further increase the feasibility of incorporating vision sensor technology in low-power and small-form-factor wearable smart sensors. Likewise, the SOA in IMU technology has reached such as point that open-source data fusion algorithms can provide accurate and precise orientation measurements [46]. These advances in the vision and IMUs create a need for novel multimodal sensor fusion algorithms to utilize these emerging possibilities to perform human motion tracking using less expensive, smaller, and less complex tracking systems.

Therefore, the key advances that this work proposes include the wearable opto-inertial, inside-out motion tracking system that relies on two external, known, points of reference embedded in the ambient environment, i.e. IR LEDs. Moreover, a novel multimodal sensor fusion algorithm for 3-D pose detection is proposed that utilises this system architecture, i.e. the inside-out opto-inertial tracker with two known points of reference in the form of IR LEDs.

4.2 System Description

This section describes the proposed system architecture and the multimodal sensor fusion algorithm that leverages its properties. First, the overall system is described in detail. It includes the hardware specifications. Subsequently, the proposed novel algorithm development for motion tracking and its use in the exercise tracking system is described.

4.2.1 *System Architecture*

Human motion tracking using wearable smart sensors requires a thoughtful consideration of many factors, especially in the context of applications that require low-power and small-form-factor. The proposed WP incorporates a monocular vision sensor, which can have negative

implications on the performance. Despite its advantages, vision sensors require a considerable amount of computational power to process multiple Frames Per Second (FPS), each with many pixels; often counted in millions, i.e. Mega Pixels (MP). The WP needs to be able to process the image frames at a relatively high frame rate; in tens of FPS. Furthermore, the type of information that needs to be extracted from the image frames has a significant impact on the complexity of the image processing algorithms used in this task. For example, a high noise floor in the images, accompanied by the complexity of the points of interest to be found, can dramatically increase the computational requirements of the system. Hence, a human motion tracking system in this context needs to consider all factors; including the software/firmware, hardware as well as the ambient environment beyond the WP.

The system proposed in this work can be broken down into two main elements, the WP and the Ambient Environment, as shown in Figure 32. These two elements are connected via an RF telecommunications link. The RF link enables an interaction between these two elements to help ensure that the system operates in its optimum conditions.

The WP incorporates a monocular vision system and an IMU to perform the inside-out tracking. It also has an MCU for data processing, power management block and an RF module. The Ambient Environment consists of and RF module with an MCU and the points of reference. This system was designed with active markers as the points of interest to be tracked. The Infrared IR LEDs are tracked by the camera in the WP, which has a matching IR filter attached to it.

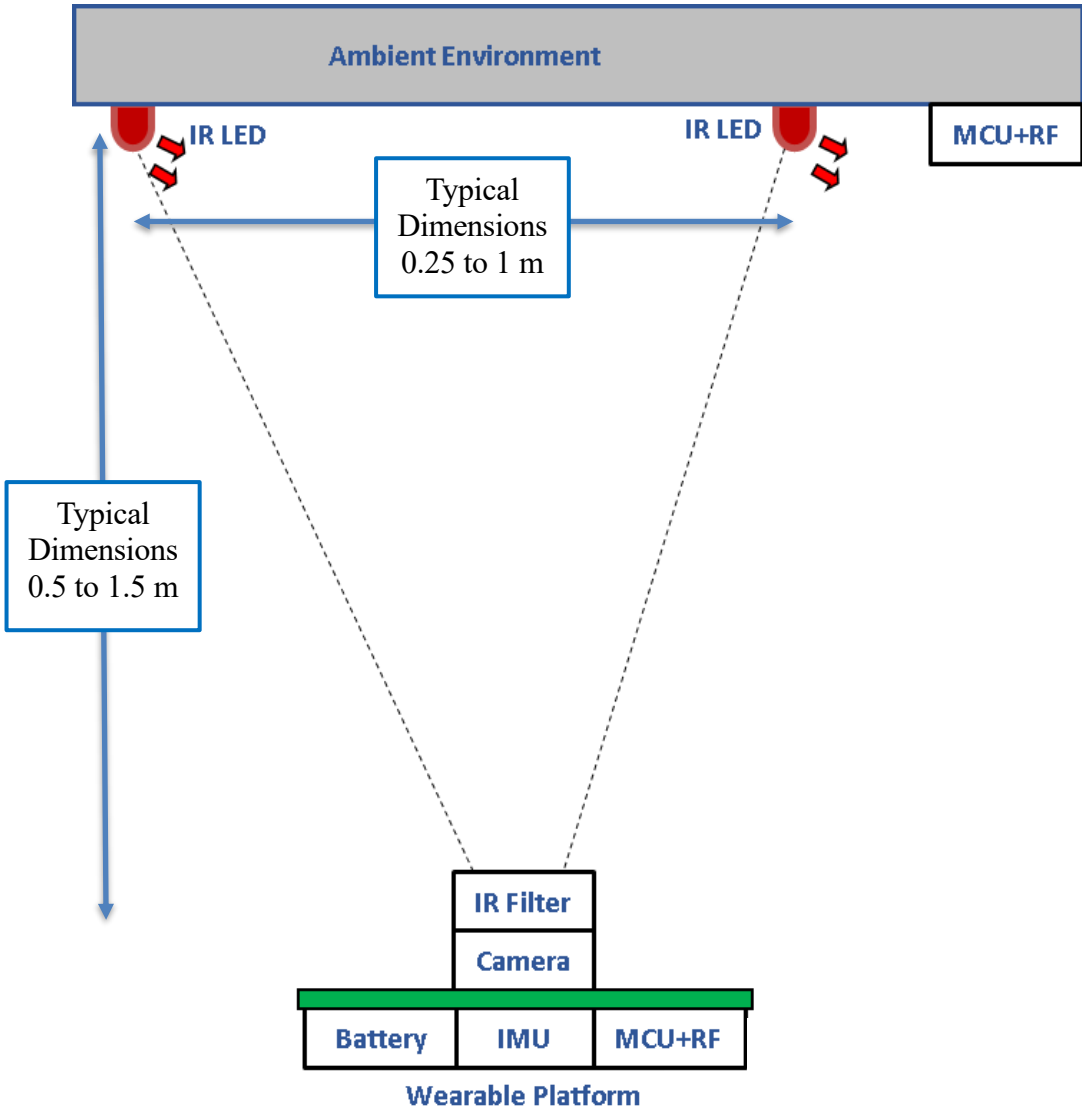


Figure 32: Generalised System Architecture

The Ambient Environment (AE) incorporates two IR LEDs and a control unit to drive them and maintain the optimum conditions for the WP. The control of the IR LEDs is carried out in a similar manner to that described in Chapter 3. The optimum conditions of the system are such that the intensities of the two IR LEDs are set so as to ensure that their pixel intensity profiles, measured by the camera, are in the optimum range in all 3-D poses of the WP; i.e. neither too high (no saturated pixels) nor too low (point peaks are not buried in the noise floor). It is important because changes in position and/or orientation of the WP cause changes in intensities and dimensions of the IR LEDs as captured on the camera's pixel array, which in turn can have a negative effect on the performance of the point detection and tracking tasks. Therefore, these changes need to be offset by controlling the intensity of the IR LEDs. In practical terms, the

intensity is maintained in the range of values between 33 % and 66 % of the maximum intensity. The MCU in the AE maintains the IR LED intensity with Pulse Width Modulation (PWM), given the control signals it receives from the WP via the RF link. The intensity range of the LEDs is mapped to the 8-bit pixel intensity range of the camera in the WP. For example, the pixel intensity of 85 on the pixel array, on the camera in WP, corresponds to approximately 33 % of the LED's intensity range. Likewise, 170 corresponds to approximately 66 % of the LED's intensity. These parameters were determined through empirical testing. This testing was motivated by the requirements associated the SLI algorithm. Therefore, the points on the pixel plane, as captured by the camera, had to be spread over between 3x3 and 6x6 pixels, as described in Chapter 3. This work included a careful component selection, i.e. the camera and the IR LEDs. The consideration of working distance ranges between the WP and IR LEDs was also important.

The work envelope, i.e. the space in which the WP can operate, of the system was designed with simplicity and scalability in mind and is driven by the potential end-use-case requirements. It defines the volume of 3-D space within which the proposed system can perform its intended motion tracking function. Since the proposed 3-D pose detection algorithm relies on two points of reference in the ambient environment, details of which will be described in the following section, the two IR LEDs must be within the FoV of the WP's camera. Also, given the fact that most of ST exercises, one of the potential target use cases, are largely stationary, the work envelope doesn't need to be large. Though, it needs to be scalable to facilitate other potential future use cases. As a result, the work envelope for the system was designed with an arbitrarily set distance between the IR LEDs, called the baseline $B = 500\ mm$. This value of B allows for meeting two objectives. Firstly, the WP can perform translation within the work envelope with a wide range of rotations, while retaining both reference points in the FoV of the camera. Secondly, the calculations in the proposed algorithm yield more accurate results if the distance between the LEDs, as captured by the camera, is relatively large, i.e. such that the two points captured in the images are far from one another, since the proposed algorithm relies on the geometries formed in the system, which is described in detail in the following section.

The size of the work envelope can be scaled, up or down, by adding additional IR LEDs separated by the baseline distance B. The multiple IR LEDs can be switched ON and OFF using the RF link; as the WP moves through the 3-D space. The size of the work envelope can also be changed by varying the value of the baseline B. However, the scope of this work is to describe the fundamental principles of this system. Thus, the use of two IR LEDs with a fixed distance B is

sufficient. The work envelope is shown in Figure 33. It is effectively a 3-D space whose boundaries are defined by the continuous thick line segments shown in the figure. Its dimensions have a twofold impact on the system. Firstly, the intensity of the LEDs can be controlled dynamically to maintain the optimum level for the camera in the WP. Secondly, both reference points remain within the FoV of the camera; with the exception for certain orientations in the boundary regions. These parameters match the requirements of one of our potential target application space especially that of certain ST exercises, such as the barbell squat. The WP, or potentially more than one WP, can be attached to the back of the athlete to track its motion during executing the squat. The information extracted from the motion trajectories captured by the WP can help determine whether the exercise if carried out correctly.

It needs to be noted that the naming conventions from robotics engineering were adopted in this work. The right-handed coordinate system was used. The origin of the global, or World, coordinate frame L_W is coincident with the location of the reference point P_0^W (read as point zero in World reference frame), as shown in Figure 33. The two IR LEDs were located $1000\ mm$ above the ground, thus placing the origin of frame L_W at that height.

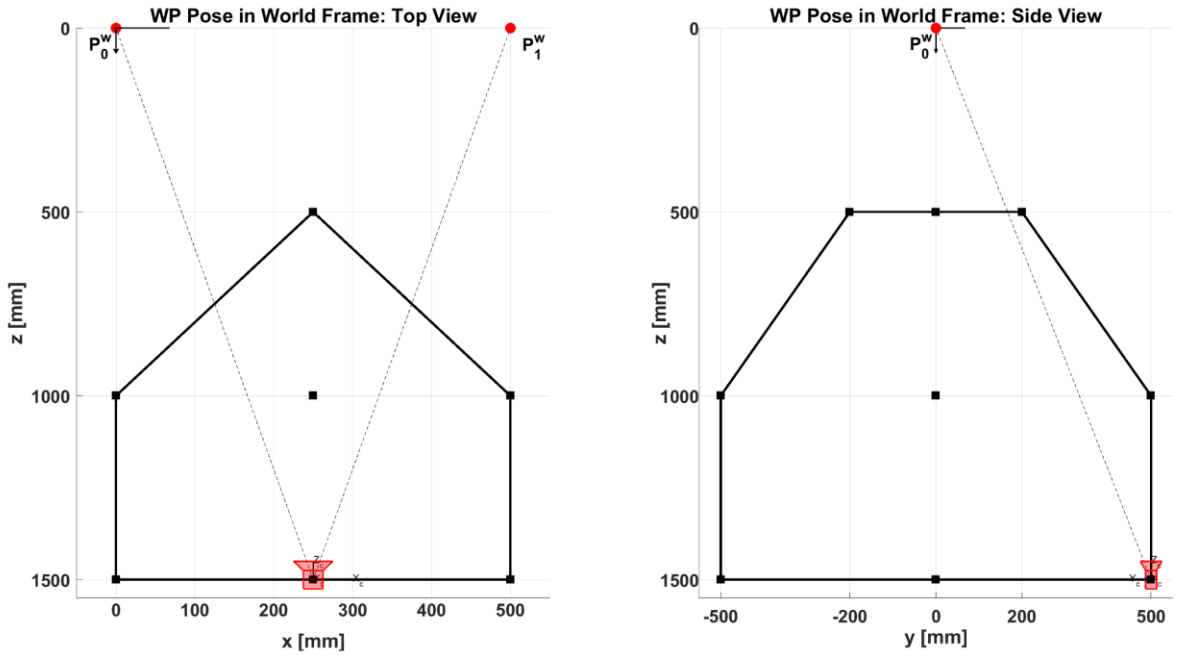


Figure 33: Wearable Platform (Represented by the camera) inside the Work Envelope (Thick Continuous Line) with Reference Points P_0 and P_1 (IR LEDs) in Camera's FoV in World Coordinate Frame

4.2.2 *3-D Pose Detection Algorithm*

The proposed sensor fusion algorithm performs the 3-D pose detection function directly using the data from the calibrated vision and IMU sensors, as shown in Figure 34. The proposed system architecture, shown previously in Figure 32, simplifies this process. At the pre-processing stage in Figure 34, the coordinates of the two points are efficiently extracted from the image frames. Firstly, we use IR light spectrum to suppress the ambient light and reduce the problem of point detection to a local-maxima detection routine. Secondly, we use our proposed novel algorithm for subpixel point detection described in Chapter 3, to reduce the resolution of the camera without compromising the accuracy [91]. In terms of the IMU, we use a SOA algorithm for accurate orientation estimation of the WP, after a minor calibration routine which is aimed at aligning WP's reference frame with World reference frame in terms of orientations. The details of this calibration are described later in this section. Back in 2011, Madgwick et al. demonstrated an efficient and accurate IMU sensor fusion algorithm for orientation calculation [46]. Our algorithm fuses the data from these two calibrated sensor modalities and calculates the 3-D pose that can be passed to the subsequent stages, e.g. a data aggregator in a local body area network, an HMD or some other type of the Human Computer Interface (HCI).

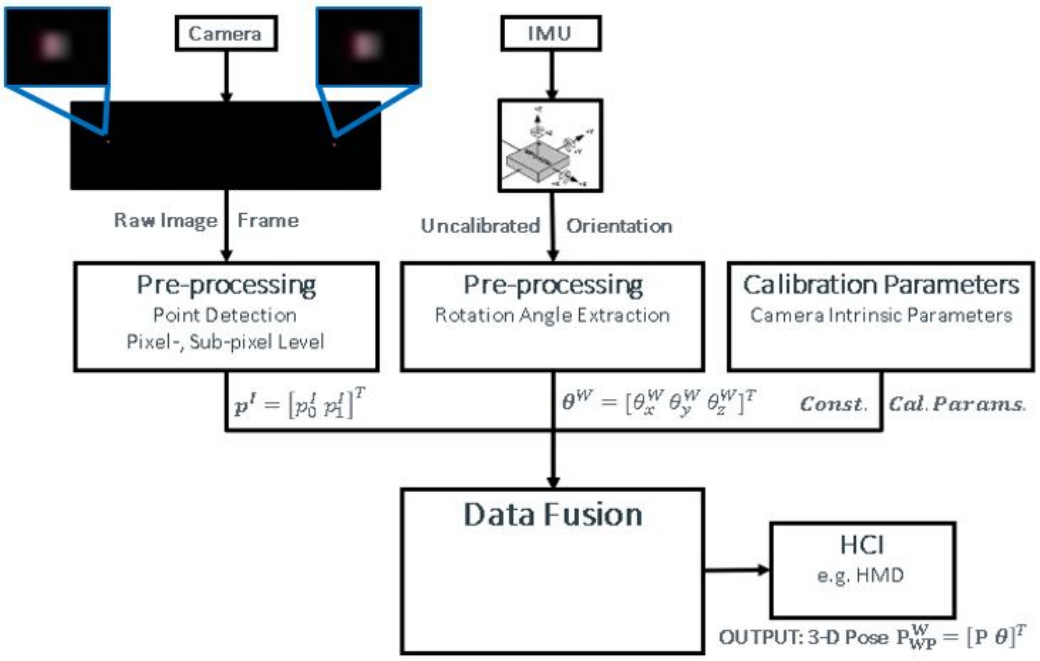


Figure 34: General Block Diagram of the Proposed Data Fusion System (Raw Input Frame Contains Two Points of Reference)

The key to the 3-D pose estimation is the use of the vision sensor. The pose of the camera can be determined by solving the PnP problem. The solution of the PnP problem involves estimating the pose of the calibrated camera given a set of known 3-D points in the world and their 2-D projections on the camera's image plane. There exist several algorithms that can solve it. The SOA methods require that the number of known points is $n \geq 3$ [18, 55]. Our proposed method takes a different approach. In our case, the number of points of reference is $n = 2$. We can determine the 3-D pose from only two reference points, because we complement the missing pieces of information with the calibrated IMU data. We complement the geometries formed by the two reference points and the camera with the rotation angles extracted from the IMU.

The Data Fusion block is where the 3-D pose is computed. It takes in three inputs: the coordinates of the two reference points extracted from the image frame, expressed in Image frame, $p^I = [p_0^I \; p_1^I]^T$, the orientation of the WP from the IMU, expressed in the World frame of reference, $\theta^W = \left[\theta_x^W \; \theta_y^W \; \theta_z^W\right]^T$, and the camera intrinsic calibration parameters.

The orientation of the WP in World frame of reference, i.e. the vector θ^W, can be obtained by transforming the IMU's output orientation to World frame of reference. If the IMU is calibrated correctly, Madgwick's algorithm returns the orientation in Earth's frame of reference L_E; as defined by Earth's magnetic and gravitational fields [46]. Therefore, the homogenous transformation matrix from Earth, L_E, to World, L_W, frame of reference, T_W^E, can be defined as one containing a the rotation matrix with the translation elements set to zero. In practice, the y-axes in frames L_W and L_E are parallel to each other, i.e. $\hat{y}^W \parallel \hat{y}^E$, and can be assumed to be pointing in the same direction, i.e. their dot product is $\hat{y}^W . \hat{y}^E = 1$. Therefore, the transform T_W^E is reduced to describing a fixed rotation about the $\hat{y}^W$- axis. This transformation is then used for transforming the orientation of the WP from L_C to L_W, as follows. The vector of rotation angles of the WP measured by the IMU, $\theta^E = \left[\theta_x^E \; \theta_y^E \; \theta_z^E\right]^T$, can be also represented as a homogenous transformation matrix from Camera, L_C, to Earth, L_E, frame of reference T_E^C; with the X-Y-Z order of rotations in the rotation elements and the translation elements set to zero [107]. Therefore, the transformation from L_C to L_W, i.e. T_W^C, is defined as shown in (14). Subsequently, the rotation angles of the orientation vector θ^W can be extracted from this equation:

$$T_W^C = T_W^E T_E^C \tag{14}$$

The intrinsic camera parameters for the specific vision sensor can be calculated via a camera calibration process, which is a commonly used in image processing tasks for determine the key properties of a camera, such as lens distortion parameters, optical centre, focal length, to name but a few [61]. The intrinsic parameters, along with the knowledge of the specific image sensor from its datasheet, such as the focal length f, pixel dimension and size and location of the optical centre, are used to transform p^I to the Camera reference frame L_C expressed in metric units; resulting in p^C. The output is the 3-D pose of the WP defined as the position and orientation in the World frame of reference as follows $P_{WP}^W = [P\ \theta]^T = \left[P_x^W\ P_y^W\ P_z^W\ \theta_x^W\ \theta_y^W\ \theta_z^W\right]^T$. The subscripts in the variables define the axis. For example, the θ_x^W is the rotation angle about the $\hat{x}$-axis in the World frame of reference. Note, the hat symbol implies the axis component unit vector, e.g. $\hat{x}^W$ means the $\hat{x}$-axis in World reference frame. The proposed data fusion algorithm computes the 3-D pose in three discrete steps, as shown in Figure 35.

- Step 1 corrects the input points p^I using the rotation angle of the WP about the $\hat{z}$-axis in World frame θ_z^W. The subsequent two steps break down the problem into two smaller tasks.
- In Step 2, the position P_x^W is computed on the $\hat{x}^W\ \hat{z}^W$-plane with θ_y^W.
- In Step 3, the position elements P_z^W and P_y^W are computed on the $\hat{y}^W\ \hat{z}^W$ −plane and with θ_x^W, to finally yield the result, i.e. the 3-D pose of the WP in the World frame of reference P_{WP}^W.

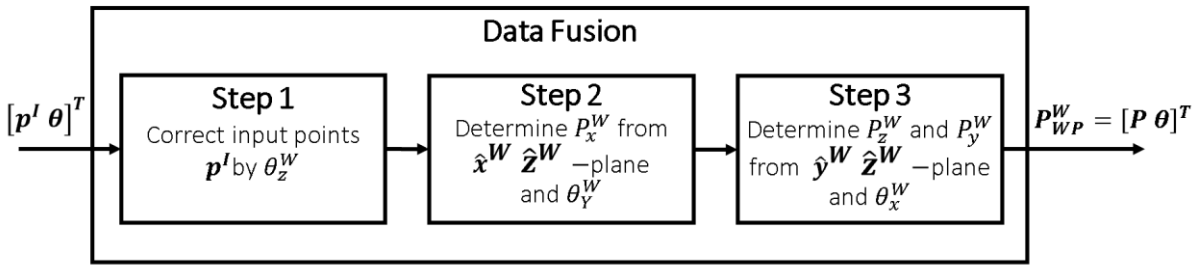


Figure 35: Block Diagram of the Proposed Data Fusion Algorithm

The three steps of the proposed algorithm, shown in Figure 35, are described in detail in the subsections below.

4.2.2.1 Step 1- Input Points Correction

The geometric model that is used in calculating the pose of the WP achieves the best results when the $\hat{x}$-axis or the $\hat{y}$-axis of the L_C, i.e. that of the WP, and L_W reference frames are parallel, or close to it. It is so because the calculations in Steps 2 and 3 of the proposed algorithm are carried out on the planes $\hat{x}^W\hat{z}^W$ and $\hat{y}^W\,\hat{z}^W$, respectively. In other words, the calculations are more accurate if the rotation matrix from L_C to L_W reference frame R_W^C is as close as possible to that defined in equation (15). This condition means that all corresponding axes are parallel; with $\hat{y}$-axes and $\hat{z}$-axes of these two reference frames pointing in opposite directions. It simplifies the geometry formed by the IR LEDs and the camera. Effectively, the line segment between points p_0^I and p_1^I extracted from the image frames needs to be parallel with the $\hat{x}$-axis of the frame L_C. However, it is not a realistic scenario. It effectively makes the WP's orientation constant, such that it directly faces the IR LEDs, with only the translation being allowed to vary. It is obviously an unacceptable condition in the context of the considered application space. Therefore, our algorithm uses a corrective step to meet this condition, or at least approximately match it.

$$R_W^C = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{15}$$

The corrective step is applied to point p^C. Whereas it would be a straightforward process in 3-D, it is more complicated in the case of the two points p^C. In the case of 3-D points the data from the calibrated IMU could be used to rotate the points. However, the translation vector of the WP P is unknown at this step. In fact, the objective of this work is to determine P.

The proposed solution to this problem takes advantage of the fact that many ST exercises are largely stationary, i.e. in one location, with a predefined body posture and range of motion. For example, a barbell squat would involve relatively little rotation and some translation if the WP was attached to the back of the exerciser. From a technical point of view, this means that the WP would face the reference points in the ambient environment. It needs to be noted that the initial rotation matrix R_W^C is the same as that defined in equation (15). Also, the rotation angles would be relatively small. Hence, our corrective step involves a two-dimensional rotation of the image

points p^C by rotation angle θ_z^C, as defined in (16). This angle is not negative, because we are correcting the orientation of the WP. The operation or rotating points p^C by θ_z^C, which is effectively θ_z^W, is an approximate equivalent of rotating the WP in the opposite direction.

$$Rot(\hat{z}^C, \theta_z^C) = \begin{bmatrix} cos\,\theta_z^C & -sin\,\theta_z^C \\ sin\,\theta_z^C & cos\,\theta_z^C \end{bmatrix} \tag{16}$$

Subsequently, the two transformed points p^C are passed to Step 2 in the algorithm.

4.2.2.2 Step 2 – Calculation of P_x^W

In this step, the position P_x^W of the WP is computed using the $\hat{x}^W\ \hat{z}^W$-plane, as shown in Figure 32. The $\hat{y}$-axis is ignored in this step, because the algorithm performs the calculation only on the $\hat{x}^W\ \hat{z}^W$-plane. The elements of the general system architecture, shown in Figure 32, directly correspond to the geometric model shown in Figure 36. The IR LEDs correspond to the points P_0^W and P_1^W while the camera is expressed as the large rectangle. To simplify the model, the IR filter and lens were assumed to be ideal elements that don't affect the system.

This model enables the calculation of the 3-D pose due to its specifically designed architecture. Firstly, the baseline B is known. Secondly, the camera's intrinsic parameters can be determined by camera calibration. The camera calibration routine can determine the key parameter of the camera that is critical in the calculations, i.e. the focal length f. Furthermore, the knowledge of these parameters, complemented with the rotation angles from the IMU, enabled us to use geometry and trigonometry to compute the pose.

The knowledge about the orientation of the WP makes it possible to use geometry to solve the problem of determining the 3-D position. The properties of similar triangles and trigonometry are particularly useful. The camera can be modelled with a simplified projection model, i.e. one in which the image plane is in front of the principal point, which is coincident with the origin of the Camera frame L_C; as opposed to being behind it. The image points p_0^I and p_1^I are the projections of their corresponding World points P_0^W and P_1^W on the camera's image plane. The two rays of light, R_L and R_R, that originate from the two World points and pass through their corresponding Image points intersect at the L_C. The rotation angles from the IMU help us form two similar triangles. The first triangle has the following vertices P_0^W, P_1^W, and L_C. The second

triangle has the following vertices p_0^I, p_1^I, and L_C. The image points p_0^I, p_1^I are transformed to the Camera frame to enable real-world-unit calculations, i.e. p_0^C, p_1^C. The proportions are achieved by making B and B' parallel.

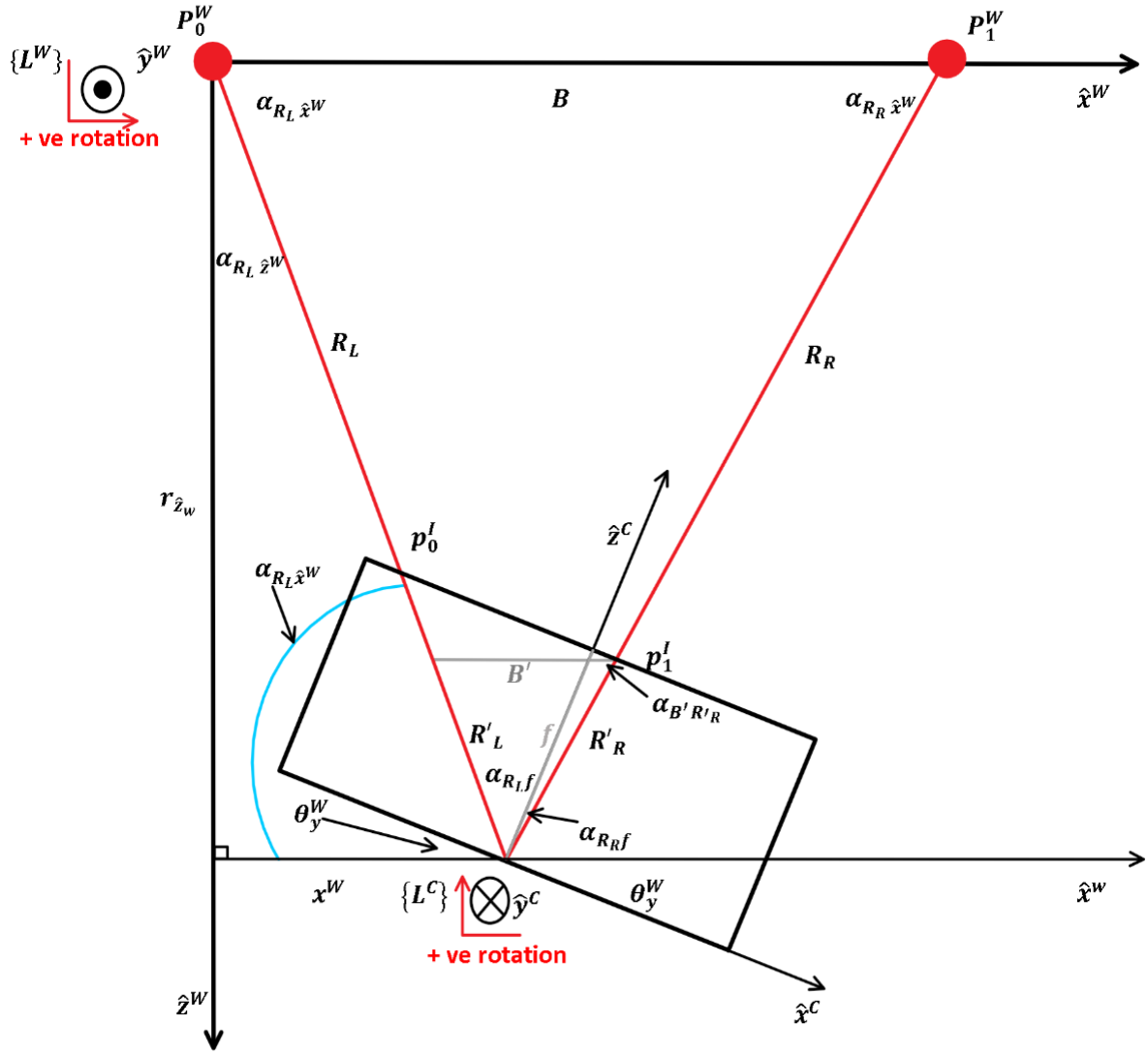


Figure 36: Geometric model of the system, x-z plane in World Coordinate Frame

The first task in this step is to compute the angles: between the left light ray R_L and the line segment of length equal to the focal length f, angle between R_R and f, angle between the R_L and f, angle between R_L and the axis $\hat{x}^W$, angle between R_L and the axis $\hat{z}^W$, angle between the

rays R_L and R_R, and the angle between B' and R'_R; defined in: (17), (18), (19), (20), (21) and (22), respectively.

$$\alpha_{R_L f} = tan^{-1}\left(\frac{p_0^C}{f}\right) \tag{17}$$

$$\alpha_{R_R f} = tan^{-1}\left(\frac{p_1^C}{f}\right) \tag{18}$$

$$\alpha_{R_L \hat{x}^W} = \frac{\pi}{2} + \alpha_{R_L f} - \theta_y^W \tag{19}$$

$$\alpha_{R_L \hat{z}^W} = \frac{\pi}{2} - \alpha_{R_L x^W} \tag{20}$$

$$\alpha_{R_L R_R} = \left(\alpha_{R_R f} - \alpha_{R_L f}\right) \tag{21}$$

$$\alpha_{B' R'_R} = \pi - \alpha_{R_L \hat{x}^W} - \alpha_{R_L R_R} \tag{22}$$

The length of the line segment R'_L is calculated with (23), which then allows us to determine the value of B' using (24), using the sine rule and transposing (25).

$$R'_L = \sqrt{f^2 + {p_0^C}^2} \tag{23}$$

$$\frac{B'}{\sin\left(\alpha_{R_L R_R}\right)} = \frac{R'_L}{\sin\left(\alpha_{B' R'_R}\right)} \tag{24}$$

$$\therefore B' = \frac{R'_L \sin(\alpha_{R_L R_R})}{\sin(\alpha_{B' R'_R})} \tag{25}$$

The properties of Similar Triangles can be used to find the length of R_L with (26) followed by (27).

$$\frac{R_L}{R'_L} = \frac{B}{B'} \tag{26}$$

$$\therefore R_L = R'_L \left(\frac{B}{B'}\right) \tag{27}$$

In the final stage, trigonometry is used to find the vlalues of the remaining two variables. The sine function is used to find x^W with (28) and (29), which is in effect equal to one of the elements of the 3-D Pose P_x^W.

$$\frac{x^W}{R_L} = sin(\alpha_{R_L \hat{z}^W}) \tag{28}$$

$$\therefore x^W = R_L sin(\alpha_{R_L \hat{z}^W}) \tag{29}$$

Finally, the value of $r_{\hat{z}_W}$ is computed using the cosine function with (30) and (31). The radius $r_{\hat{z}_W}$ is required in the computations in Step 3.

$$\frac{r_{\hat{z}^W}}{R_L} = cos(\alpha_{R_L \hat{z}^W}) \tag{30}$$

$$\therefore r_{\hat{z}^W} = R_L cos(\alpha_{R_L \hat{z}^W}) \tag{31}$$

4.2.2.3 *Step 3 - Calculation of P_z^W and P_y^W*

The remaining two unknown variables are computed in this step, i.e. the y^W and z^W. The y^W and z^W correspond to the P_y^W and P_z^W elements of the P_{WP}^W vector, respectively. The computations are carried out on the $\hat{y}^W\ \hat{z}^W$ −plane. The corrective rotation, that was applied in Step 1, allows us assume that the axes of the frames L_C and L_W are approximately aligned with the rotation transformation R_W^C close to that defined in equation (15), i.e. the x-y planes defined by axes in Camera and World frames are parallel, i.e. $\hat{x}^c$- $\hat{y}^c||\hat{x}^W$- $\hat{y}^W$. As in the previous step, the system setup allows us to use trigonometry to determine the missing pieces of information. The geometric model of the system is shown in Figure 37. It is effectively the side-view of the system. The calculations use three inputs. Given the corrections described in Step 1, the line segment formed by the image point vector p^I is effectively parallel to $\hat{x}^W$, correct to approximately within 1 deg. The mid-point between these two points p_{01}^I is used; specifically, the vertical coordinate on the image plane. As in the previous step, the p_{01}^I is transformed to p_{01}^C for calculations in real-world-units. Also, the $\hat{x}^W$-axis is ignored in this step.

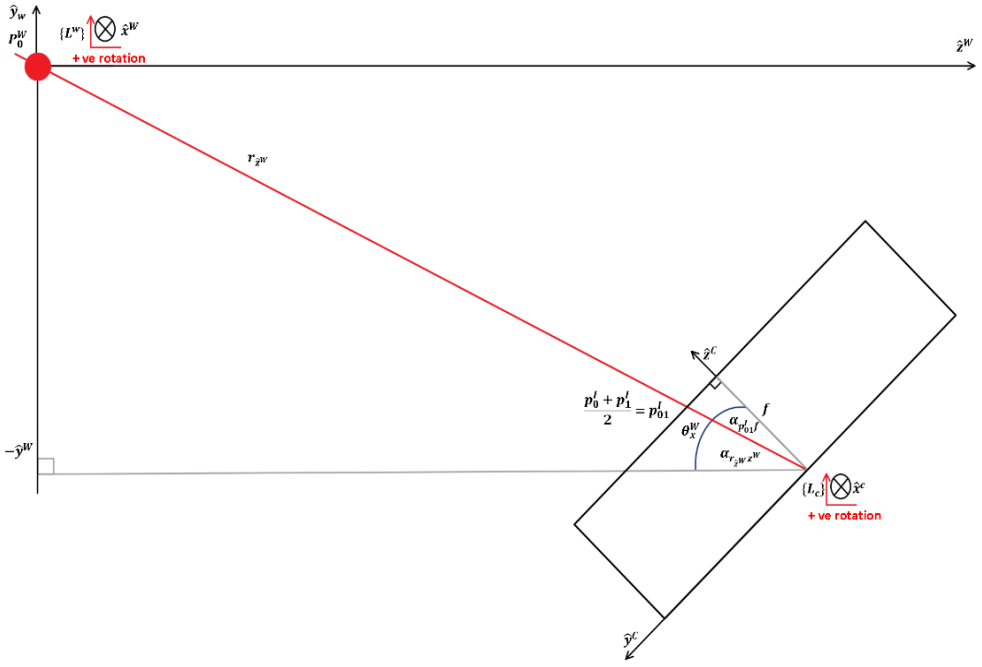


Figure 37: Geometric model of the system, y-z plane in World Coordinate Frame

The angle $\alpha_{p_{01f}^C}$ is found using the right-angled triangle with vertices at: the intersection of $\hat{z}^C$ with image plane, the mid-point p_{01}^C, and the origin L_C. Thus, the inverse tangent of of the ration of the p_{01}^C to the focal length f is equal to this angle, as defined in (32). The angle between the $\hat{z}^W$-axis light-ray R_L, whose length is $r_{\hat{z}}$w, is found by correcting $\alpha_{p_{01f}^C}$ by θ_x^W, as shown in (33). Finally the remaining unknowns z_W and y_W are found using cosine and the negative sine functions of $\alpha_{r_{\hat{z}W}\hat{z}W}$, scaled by $r_{\hat{z}W}$, defined in (34) and (35), respectively.

$$\alpha_{p_{01f}^C} = tan^{-1}\left(\frac{p_{01}^C}{f}\right) \tag{32}$$

$$\alpha_{r_{\hat{z}W}\hat{z}W} = \theta_x^W - \alpha_{p_{01f}^C} \tag{33}$$

$$z_w = r_{\hat{z}W} cos\left(\alpha_{r_{\hat{z}_w}\hat{z}W}\right) \tag{34}$$

$$y_w = -r_{\hat{z}W} sin\left(\alpha_{r_{\hat{z}_w}\hat{z}W}\right) \tag{35}$$

At this point the 3-D Pose is computed. The elements of the pose vector are as follows: $P_{WP}^W = [P\ \theta]^T = \left[P_x^W\ P_y^W\ P_z^W\ \theta_x^W\ \theta_y^W\ \theta_z^W\right]^T = \left[x^W\ y^W\ z^W\ \theta_x^C\ -\theta_y^C\ -\theta_z^C\right]^T$. The orientation angles θ, measured by the IMU, determine the orientation of the WP. The orientation of the WP in Word and Camera frame are the same, with the exception for the signs of some of its elements; since WP faces the IR LEDs, and the rotation matrix R_W^C is assumed to be relatively close to that defined in (15).

4.3 System Modelling and Simulations

Prior to the implementation stage, the proposed system, along with the proposed sensor fusion algorithm, was modelled and evaluated in simulated conditions. The objective of this task was twofold. Firstly, the system's performance was to be simulated in a number of scenarios.

Secondly, the impact of various noise levels originating from uncertainties in point detection and orientation estimation processes was to be determined. The proposed system was modelled and evaluated in MATLAB®.

One of the key elements in modelling the system is the camera to be used in the data capture. To be able to simulate it in a realistic way, the camera had to be carefully modelled. The locations of the two input points of reference p_0^I and p_1^I, captured by the camera, as visualised in Figure 34, had to closely correspond to their respective locations in the World frame, as shown in Figure 33. This correspondence was critical in achieving the ability to compare the results calculated by the proposed system to the real-world position and orientation of the WP. The pinhole camera model is commonly used to map 3-D World points to 2-D Image points [108]. In this work, we used a MATLAB® implementation of this model developed by Zachary Taylor [109]. It was used for projecting 3-D points onto a 2-D image plane using camera calibration parameters, the 3-D coordinates of the two reference points P^W and the extrinsic matrix. The camera calibration parameters were obtained from the same camera module that was used in the experimental work (described in Section 4.4). Likewise, the focal length f, which was required by the proposed algorithm, was obtained from the intrinsic matrix. The extrinsic matrix is a transform that describes pose of the WP in the World frame of reference. Thus, the input position and orientation of the WP in the World frame of reference was encoded in this transform matrix and passed to the function that projected the two 3-D reference points and output the 2-D image points. The two Image points were subsequently used as one of the two inputs to the proposed data fusion algorithm. The second input was the orientation vector θ^w, which was also used in constructing the extrinsic matrix.

4.3.1 *Simulated Scenarios*

The proposed system was evaluated in several scenarios. In each case, $N > 5000$ appropriate inputs were generated and passed to the data fusion algorithm. The following scenarios were used in this process.

4.3.1.1 Scenario 1 - Linear motion – along $\hat{x}\hat{y}\hat{z} - axes$

In this scenario, the WP moved on a straight line across the Work Envelope along all three axes, i.e. $\hat{x} - \hat{y} - \hat{z}$ in the World frame of reference. The translation along the axes was as follows: $x^W \in < 150, 350 >$ mm, $y^W \in < -250, 250 >$ mm, $z^W \in < 1000, 1500 >$ mm. The orientation vector was set to $\theta^W = [0, 0, 0]^T$ deg, and it did not vary.

4.3.1.2 Scenario 2 – Uniform Random

In this scenario, the proposed system was evaluated under the most challenging conditions because both position and orientation of the WP varied at random using random number generator with uniform probability distribution. All elements of the pose vector of the WP were varied simultaneously. The range of possible positions and orientations were set such that the system was evaluated under all possible poses, including the extreme ones near the edges of the Work Envelope. The position and orientation ranges were set as follows: $x^W \in < 0, 250 > mm$, $y^W \in < 0, 250 > mm$, $Z^W \in < 1000, 1500 >$ mm, $\theta_x^W \in < 0, -10 > deg$, $\theta_y^W \in < 0, -10 > deg$, $\theta_z^W \in < 0, 10 > deg$. Although the range of positions covers only 25 % of the Work envelope for $z^w \in < 500, 1500 > mm$, it is safe to expect similar performance across the remaining volume in this range of z^W as it is a symmetrical system. It needs to be noted, that a check was performed for each pose in this scenario to ensure that both points of reference were present in camera's FoV, which was the prerequisite for the proposed data fusion algorithm to work. This condition was possible for such poses that $z^W \in < 500,\ 1000 >$ mm and the magnitude of the other elements of the pose vector of the WP were close to their maximum values in their respective ranges.

4.3.1.3 Scenario 3 – Linear motion – along $\hat{y} - axis$

In this scenario, the WP moved on a straight line across the Work Envelope along all the $\hat{y} -$ axis in World frame of reference. The translation was as follows: $x^W = 250$ mm, $y^W \in < -500, 300 >$ mm, $z^W = 1400$ mm. The orientation vector was set to $\theta^W = [0, 0, 0]^T$ deg, and it did not vary.

This scenario was of most interest to this work. It was designed to simulate the pattern of motion involved in the barbell squat carried out when using the correct technique. It was assumed the WP was attached to the back of the person executing the exercise, e.g. under the bar, between upper and lower back. In this case, there would not be much rotation expected about any axis [110]. The motion would be largely vertical with full range of motion, i.e. parallel squat, with little lateral hip shift or trunk lean [111-113].

4.3.2 *Error Analysis - Point and IMU Noise*

The proposed data fusion algorithm is susceptible to noise that is expected to be present in the input position and orientation vectors, p^I and θ^W, respectively. The individual sources of error

as well as their magnitude have a negative impact on the system. This subsection describes the process of quantifying the noise and its impact.

The point noise, i.e. error in the coordinates of the image points in the p^I vector, may originate from several sources. One of the most common causes for point noise are the imperfections in the camera's optical capture system, which were not sufficiently rectified by the camera calibration process. For example, the lens distortions may significantly alter the coordinates of points on the image plane; especially at larger distances between those points and the optical centre on the image plane [114]. The accuracy of point detection algorithms may also be affected if the angle between the optical axis of the camera and the line segment between its optical centre and the point of interest increases. Under these conditions, the shape of IR LED may resemble an ellipsoid on the pixel array, instead of a circle. The level of point noise may be measured in pixels. Its magnitude generally depends on the pixel resolution of the camera and where on the image plane the points were captured. The angle of the camera, relative to the given point, during image capture plays a role, too. Several empirical tests were carried out to determine the maximum level of point noise using the same camera module as that used in the calibration and experimental work (described in Section 4.4). The tests showed that the point noise was generally bounded to 10 pixels. As a result, point noise was modelled as a Gaussian noise distribution $\mathcal{N}_P(\mu_P, \sigma_P)$ with mean μ_P set to the noise-free input vector p^I for the given scenario and maximum standard deviation σ_P, thus resulting in p^I containing the added point noise. The maximum standard deviation was set to $\sigma_P = 10\ pixels$.

The IMU noise considered in this work was defined as the error in orientation angles of the WP, i.e. the vector θ^W. This noise may have numerous sources, ranging from poor IMU calibration to suboptimal configuration or the sensor fusion algorithm. Nevertheless, the error in orientation estimation, computed by sensor fusion algorithms, is generally bounded to $1\ deg$, [46]. Similarly to the point noise, the IMU noise was modelled with a Gaussian noise distribution $\mathcal{N}_{IMU}(\mu_{IMU}, \sigma_{IMU})$ with the mean μ_{IMU} being set to the noise-free input vector θ^W for the given scenario and the standard deviation σ_{IMU}, thus resulting in θ^W containing the added IMU noise. The maximum standard deviation was set to $\sigma_{IMU} = 1\ deg$.

The performance of the proposed system was evaluated by subjecting it to both noise types in each of the simulated scenarios. The level of noise was increased incrementally. In each scenario, the system was subjected to five different levels of noise, which was defined as a vector $\mathcal{N}_i = [\sigma_{Pi};\ \sigma_{IMUi}]$, where $\sigma_{Pi} = [0\ 2.5, 5, 7.5, 10;]pixels$ and $\sigma_{IMUi} =$

$[\ 0\ 0.25, 0.5, 0.75, 1;]\ deg$. At each level of noise $\mathcal{N}_i$, three different combinations of this noise were applied to the system: $\mathcal{N}_{IMU}$ only, $\mathcal{N}_P$ only, both $\mathcal{N}_{IMU}$ and $\mathcal{N}_P$. Thus, the individual and combined impact of noise could be examined. Note, the case with no added noise was examined at $i = 0$, i.e. $\mathcal{N}_0 = [0;\ 0;]$.

4.3.3 *Error Analysis Process*

The main performance metric was the RMSE, as defined in (36). The algorithm's output is defined as d_i and the corresponding reference values as $\widehat{d}_\iota$ over all N measurements in this equation.

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{d}_\iota - d_i\right)^2} \tag{36}$$

The RMSE was computed for each parameter as follows: simulated scenario, noise level and noise source combination. The RMSE was determined for each position element of vector P_{WP}^{W}, as well as the combined error over all three axes. The results for scenarios1, 2 and 3 are shown in Figure 38, Figure 39 and Figure 40, respectively. It can be seen that the RMSE increased in all three scenarios with the increase in noise level $\mathcal{N}_i$. The IMU noise $\mathcal{N}_{IMU}$, in most cases, has a greater impact on the RMSE then the point noise $\mathcal{N}_P$. Due to the random distribution of both noise sources, $\mathcal{N}_{IMU}$ and $\mathcal{N}_P$, the RMSE was lower than the sum of the individual RMSE values when both noise sources were applied to the system, i.e. both $\mathcal{N}_{IMU}$ and $\mathcal{N}_P$; as compared to the conditions with noise sources applied separately, i.e. either $\mathcal{N}_{IMU}$ or $\mathcal{N}_P$.

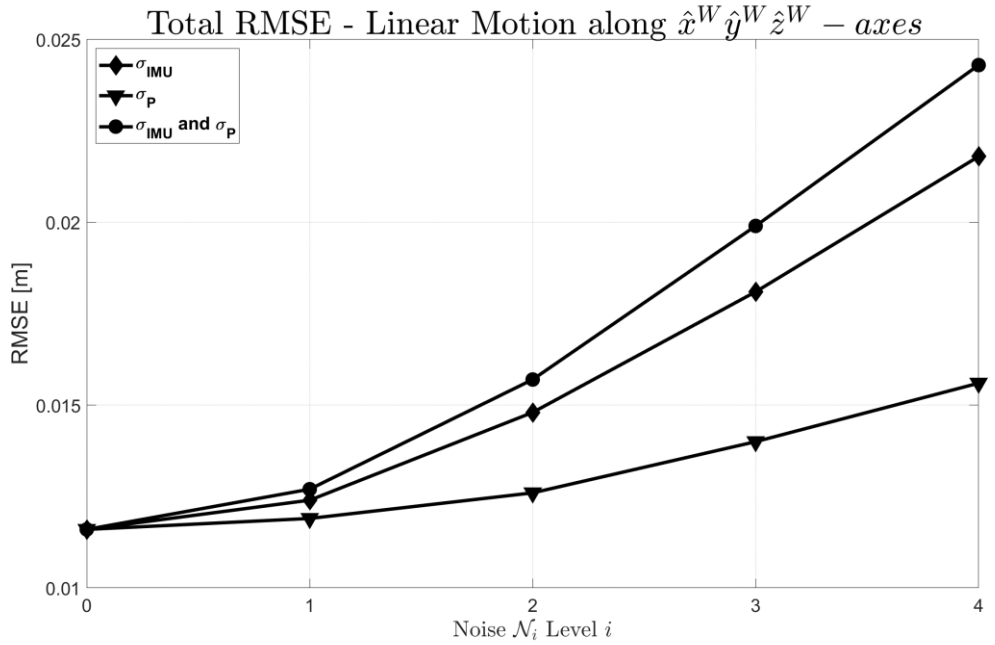


Figure 38: RMSE in Scenario 1 - Linear Motion along $\hat{x}^W \hat{y}^W \hat{z}^W$- axes for different levels of noise $\mathcal{N}_i$

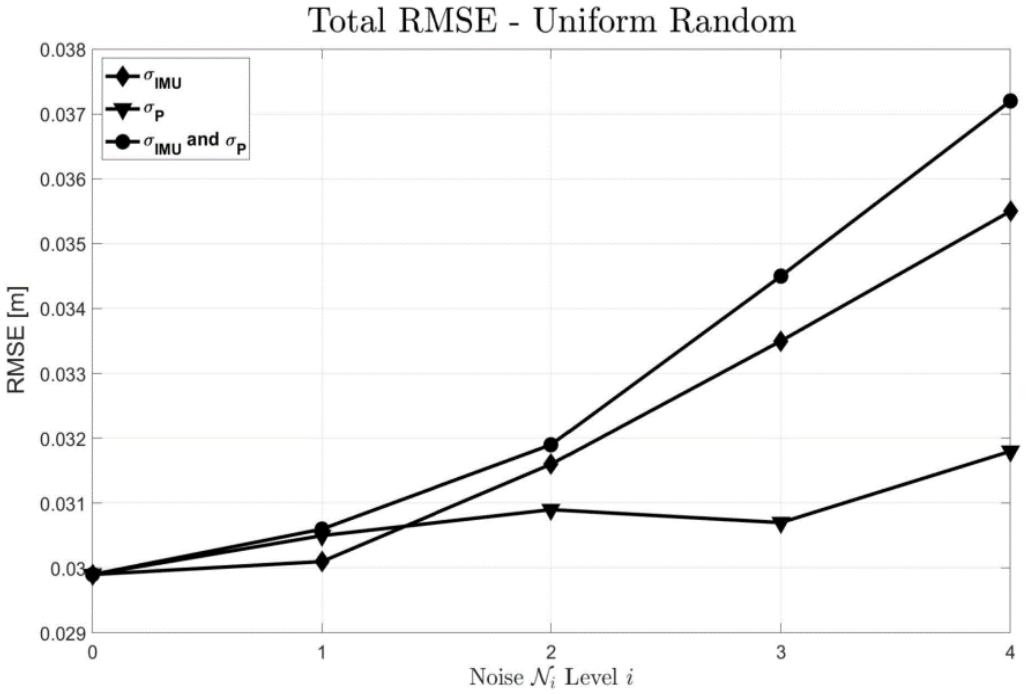


Figure 39: RMSE in Scenario 2 – Uniform Random for different levels of noise $\mathcal{N}_i$

The system achieved the lowest RMSE in scenario 1. Although, the position of WP varied across all three axes, the range of motion was relatively small, thus avoiding the unfavourable conditions. On the other hand, scenario 2 was the most challenging since both position and orientation of the WP varied at random. It was designed to determine the performance in the most adverse conditions under which it the proposed system could still perform without failing. The system would fail if the any one of the two reference points was outside of camera's FoV, or the intensity of the IR LEDs was too low for the camera to capture. As a result, the RMSE was the highest in this case. Nevertheless, the RMSE was not significantly higher in this scenario, i.e. scenario 2 which used uniformly randomly generated inputs, as compared to scenario 1.

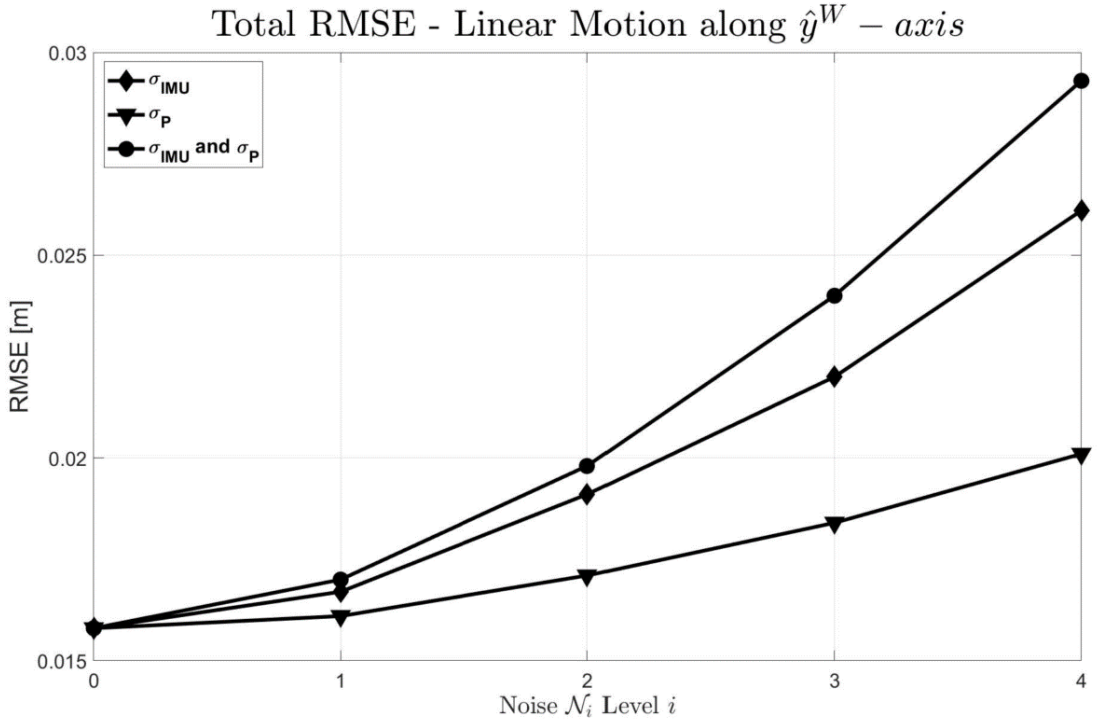


Figure 40: RMSE in Scenario 3 - Linear Motion along $\hat{y}^W$ - axis for different levels of noise $\mathcal{N}_i$

The total RMSE can be broken down into individual components, i.e. the errors in the three position elements P_x^W, P_y^W, P_z^W of the pose vector P_{WP}^W. The analysis can show that the RMSE was not equally distributed across these three position elements, and it depended on the noise level $\mathcal{N}_i$. The RMSE on all axes for different levels of noise $\mathcal{N}_i$ is shown in Table 11. The RMSE in P_z^W was the largest component of the total RMSE. Its value was the closest to the overall RMSE whereas RMSE in P_x^W and P_y^W was significantly lower. The RMSE in P_y^W was

much higher than that in P_x^W at low values of $\mathcal{N}_i$.The difference in RMSE between P_x^W and P_y^W decreased with increasing values of noise $\mathcal{N}_i$. The RMSE in P_y^W was approximately 50 % lower than that in P_z^W at low noise level $\mathcal{N}_i$ and approached as the noise increased.

The visual representation of position computation in scenario 3 is shown in Figure 41. This simulation was executed with noise level $\mathcal{N}_1 = [\sigma_{\mathrm{P1}};\ \sigma_{\mathrm{IMU1}};] = [2.5;\ 0.25;]$ [pixel; deg] to show the impact of added noise. This figure shows visually why the RMSE was lower for P_x^W and P_y^W as compared to P_z^W, as shown in Table 11. Whereas P_z^W deviated away from its reference position as the WP approached the minimum and maximum values of y^W, P_x^W and P_y^W tended to remain close to their corresponding reference values. Thus, the RMSE in P_x^W and P_y^W was relatively low and uniform as compared to RMSE in P_z^W, which was higher and increased near the minimum and maximum values of y^W. For instance, RMSE in P_x^W and P_y^W was lower than RMSE in P_z^W by a factor of approximately two for $\mathcal{N}_2$, as shown in Table 11.

The expected and acceptable performance of the system is based on the nature of the motion it is to track. An example of a particular ST exercise that this work uses as one of the demonstrator scenarios is the barbell squat. This exercise involved compound movements and often significant weights, which increases the risk and seriousness of a potential injury. For example, the lumbar spine can be at a high risk of injury if the forward trunk lean is too high while executing the squat [113, 115]. The risk of knee and hip injury significantly increase if the squats are too deep and/or the lateral hip shift occurs [110-112, 116]. In terms of accuracy, the proposed tracking system should be sufficiently accurate to reliably track the motion in this exercise for example. The specific quantitative requirements or recommendations, as to the permitted error level, are not found in the existing literature. This is because the exercise assessments are generally carried out by the coaches in a subjective manner on an individual basis following general guidelines. However, an approximate requirement for error in position tracking in the barbell squat can be estimated, based on the ranges of motion involved in this exercise. Some of the key parameters used in ensuring that the squat is executed correctly can be used as the basis for establishing this requirement. For example, the vertical range of motion in a squat carried out by an average adult individual may vary between approximately 0.5 m and 1 m, which is used in measuring the squat's depth. Likewise, the forward trunk lean can be measured by tracking the position and orientation of the line segment between two points on the back, i.e. a 3-D vector's endpoint is on the upper back, below the barbell, and origin in the lower back, on the sacral section of the spine. While the magnitude of this vector would not

vary significantly, the values of its individual components would; especially those along the vertical and the forward-facing horizontal components, i.e. the y and z, respectively. The distance between these two points on the adult athlete's back can be assumed to be approximately 0.5 m. The angle between this vector and the floor can vary between 45 degrees and 90 degrees [117]. Therefore, the values of this vector's components y and z would vary by up to approximately 35 cm. In the case of lateral hip shift, the range of motion would be smaller. It would normally reach up to a half the distance between the two feet, i.e. approximately 30 cm for an adult. Therefore, the error in position tracking of the WP would be expected to remain at single-centimetre level for this application scenario. However, the accuracy in positional tracking is not the only consideration in this application space. The motion tracking system should also be affordable and easy to setup and use. Hence, the right balance between these two factors is desired.

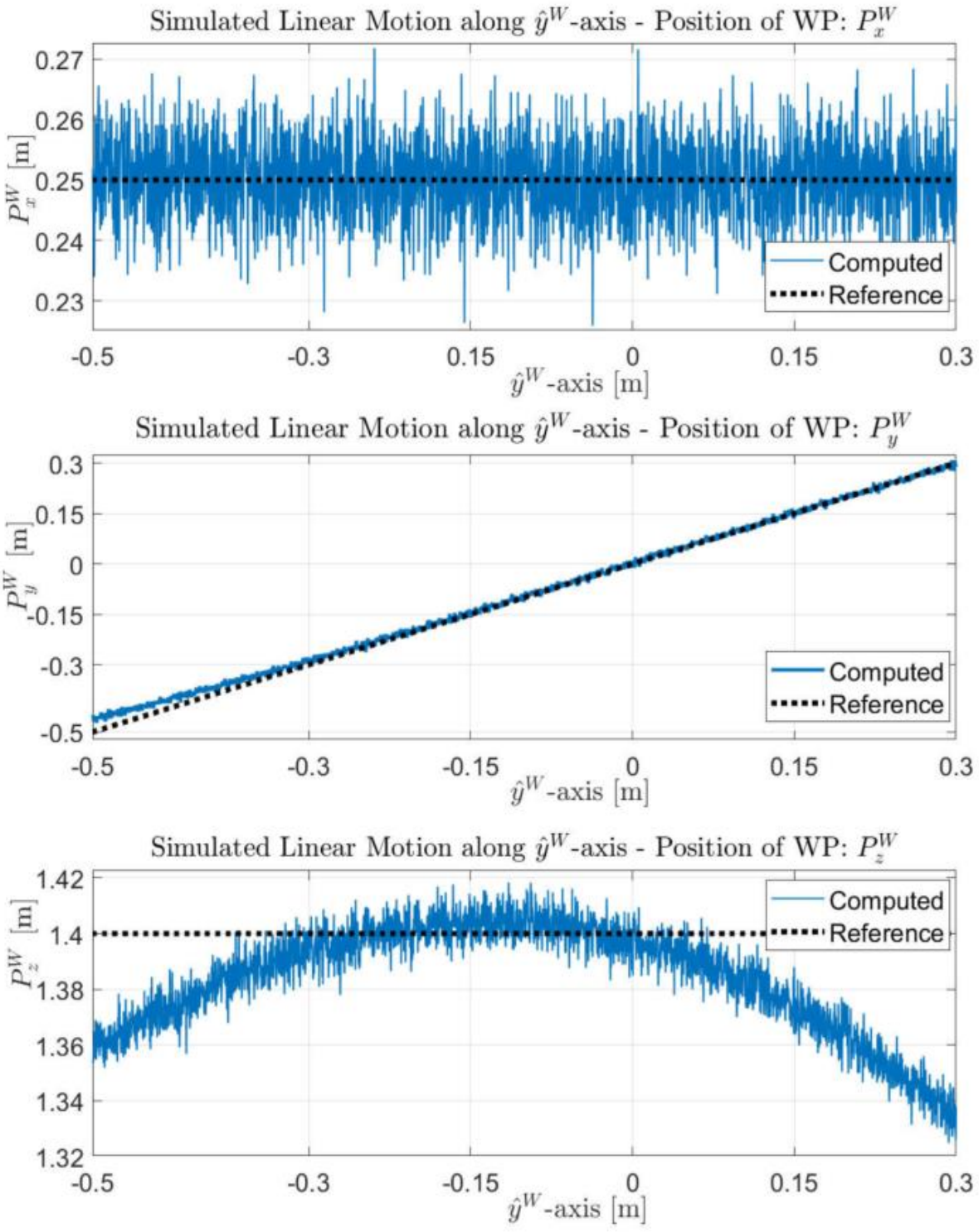


Figure 41: Simulated Position of the WP in Linear Motion along $\hat{y}^W$-axis with Added Noise $\mathcal{N}_1 = [\sigma_{P1};\ \sigma_{IMU1};] = [2.5; 0.25;]\ [pixel;\ deg;]$; Scenario 3

TABLE 11: RMSE OF ELEMENTS OF POSE VECTOR P_{WP}^{W} FOR DIFFERENT VALUES OF NOISE $\mathcal{N}_i$; SCENARIO 2

$\mathcal{N}_i - i$	RMSE P_x^W [m]	RMSE P_y^W [m]	RMSE P_z^W [m]	Total RMSE[m]
0	0.0001	0.0126	0.0243	0.0158
1	0.0065	0.0140	0.0250	0.0170
2	0.0128	0.0177	0.0265	0.0198
3	0.0195	0.0231	0.0288	0.0240
4	0.0263	0.0287	0.0324	0.0293

4.4 Experimental Validation

After modelling and simulation, the proposed system was validated experimentally. The validation process was carried out in two cases, i.e. static and mobile. In the static case, the system was validated in a similar way to that in the simulated scenario 2, i.e. the uniform random scenario. The mobile case closely resembled scenario 3, i.e. the linear motion along the $\hat{y}^W$ axis.

4.4.1 *Static Case - Experimental Setup*

The experimental setup corresponded to the general system diagram, shown in Figure 32, and the work envelope, shown in Figure 33. The complete implementation of the experimental setup is described in in Figure 42 and Figure 43. The WP was implemented using the Microsoft® Surface Pro 4 tablet computer with MATLAB® development environment installed on it. This computing platform was selected due to its portability while being a fully-featured computer. Furthermore, it had the built-in OV8865 camera module, which is a low-power camera module, designed for mobile applications. It also featured an MCU unit with a Bluetooth Low Energy (BLE) for control of the IR LEDs. An IMU, the MPU9250 from TDK InvenSense, was also added to support future functionalities. Additionally, an IR Filter was attached to the camera [94], whose transmittance properties matched the IR LEDs [96], as shown in Figure 43 (b). The WP was housed in a dedicated, 3D printed, holder that was mounted on a high-quality camera tripod. The Manfrotto MN755XB aluminium camera tripod with levelling ball with Manfrotto 410 Junior geared head were used in the experiments [118, 119]. The reference pose was measured using a digital protractor (accurate to 0.1 degree) and a laser distance meter (accurate to 1 mm) [120, 121], as shown in Figure 43 (a).

Figure 42: Experimental Setup – Static Case - Side-View

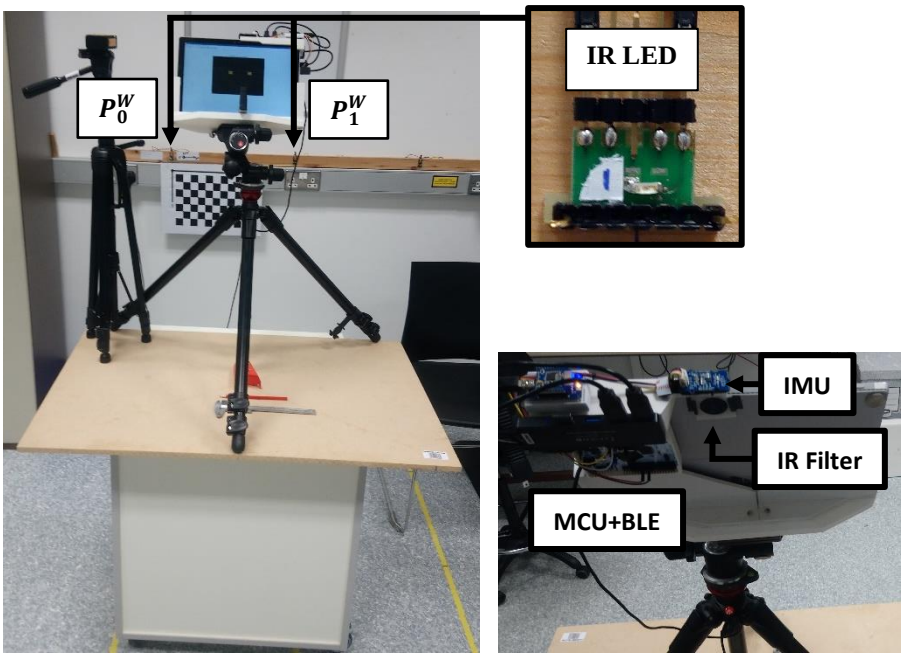


Figure 43: Experimental Setup – Static Case - (a) Front-View, (b) Rear-View

4.4.2 *Static Case - Experimental Data Acquisition*

The input dataset was acquired with the experimental setup described in the previous section. Prior to the acquisition, at each test position the intensities of the IR LEDs were set such that their perceived intensities I on the input image frame's matrix were within the following interval $I(x,y) \in < 63{,}76 >$, which was the optimum intensity for this experimental setup for our subpixel point detection algorithm [91]. Once this condition was met for the given test position, the raw input image was acquired. This process was repeated for each test position in the work envelope marked with square markers in Figure 33, except for those at $x^W > 250$ mm. Due to the symmetry along the $\hat{z}^W - axis$, at $x^W = 250$ mm, it was sufficient to consider only the work envelope with $0 \leq x^W \leq 250$ mm and $0 \leq y^W \leq 500$ mm ($0 \leq y^W \leq 200$ mm at $z^W = 500$ mm). A set of ten test positions was selected within this work envelope, with an emphasis on ensuring that all key positions along the external border were included. For each test position, an input image was acquired for the all orientations, as listed in Table 12.

TABLE 12: ORIENTATIONS FOR EACH EXPERIMENTAL TEST POSITION

Orientation Index	θ_x^W $[deg]$	θ_y^W $[deg]$	θ_z^W $[deg]$
0	0	0	0
1	-15	0	0
2	-30	0	0
3	0	-15	0
4	0	-30	0
5	0	0	15
6	0	0	30

It should be noted that certain poses at some test positions had to be excluded from the validation. The cases where one or both points of reference were beyond the FoV of the camera invalidated the input frame. The camera could not capture both points of reference at certain test positions when combined orientation angle was high, e.g. 30 degrees.

Subsequently, the raw input image frames, along with the corresponding orientation angles, were passed to the sensor fusion algorithm.

4.4.3 *Static Case - Results*

The proposed system was evaluated in the experimental laboratory environment described in section 4.4.1. It was experimentally evaluated using the same metric as that was used in simulations, i.e. the RMSE. It measured the error in the position estimation along the three axes of the World frame of reference: $\hat{x}^W$, $\hat{y}^W$, and $\hat{z}^W$. The overall RMSE over all three axes combined was also determined; referred to as the Total RMSE, which was the most important metric. RMSE was computed over $N > 1000$ measurements. The results are shown in Table 13.

TABLE 13: STATIC CASE - EXPERIMENTAL RESULTS

	$P_x^W[m]$	P_y^W $[m]$	$P_z^W[m]$	**Total [m]**
RMSE	0.0174	0.0367	0.0489	0.0367

The RMSE measurement across the individual axes revealed which position elements of the pose were more susceptible to error due to noise and the way the pose was computed by the proposed novel algorithm. It largely confirmed the pattern of noise distribution on the three axes that was present in the simulations. While the position along the $\hat{x}^W$ - axis was most accurate, the

calculation of the position along the $\hat{z}^W$-axis had the highest RMSE. These results, to some extent, correspond to the simulated scenario 2, i.e. the Uniform Random. Although this scenario did not simulate a static case, the positions and orientations of the WP were similar in both cases.

4.4.4 *Mobile Case – Experimental Setup*

An experimental setup was designed to validate the performance of the proposed system in a mobile case, as shown in Figure 44. The setup was similar to that used in the static case shown in Figure 42. It differed in that the WP was mounted on a motorised mobile track slider system. This enabled the WP to move on a vertical trajectory, along the $\hat{y}^W$-axis in a controlled manner, thus closely resembling the simulated scenario 3, as described in section 4.3.1.3, which was the main aim of this experiment. Therefore, the position and orientation and range of motion of the WP were the same as those in the simulated scenario 3. The objective of this experiment was twofold. Firstly, the RMSE was to be determined across the range of $\hat{y}^W$. Secondly, the repeatability of the performance of the proposed system was to be determined. To this end, the WP traversed the distance between $\hat{y}^W_{min}$ and $\hat{y}^W_{max}$ twenty times, i.e. it performed ten $\hat{y}^W_{min}$-to-$\hat{y}^W_{max}$-$\hat{y}^W_{max}$-to-$\hat{y}^W_{min}$ cycles.

The track slider was based on the 80 cm version of the Neewer camera slider rail, which was customised as follows for this specific experiment [122]. The slider rail was fitted with a 6-mm-wide T-belt that was connected to the Nema 17 stepper motor via matching 20-tooth pulley wheels [123]. The TB6600 stepper motor was used as the driver for the motor [124]. A Raspberry Pi® computer, Python™ programming environment and Secure Shell connection were used to control motion of the WP from a separate computer. Motion of the WP was controlled with an open-loop motor control system with a trapezoidal velocity profile. The acceleration and deceleration ramps of the velocity profile were set so as to ensure a smooth motion at the inflection points of the WP's motion trajectory, i.e. minimum, $\hat{y}^W_{min}$, and maximum, $\hat{y}^W_{max}$, values of $\hat{y}^W$. The maximum velocity was set such that the WP could acquire a sufficient amount of input frames to produce statistically significant results. The frame rate of the WP was 30 FPS. The time the WP required to traverse the distance between $\hat{y}^W_{min}$ and $\hat{y}^W_{max}$, i.e. half a cycle was $\frac{T}{2} = 17$ seconds, where T was the period of one cycle. Given ten up-down motion cycles, the WP acquired at least 5100 input frames, which was comparable to N samples in simulations. The configuration of the IR LEDS during the data acquisition process was the same as that in the static case, described previously in section 4.4.1.

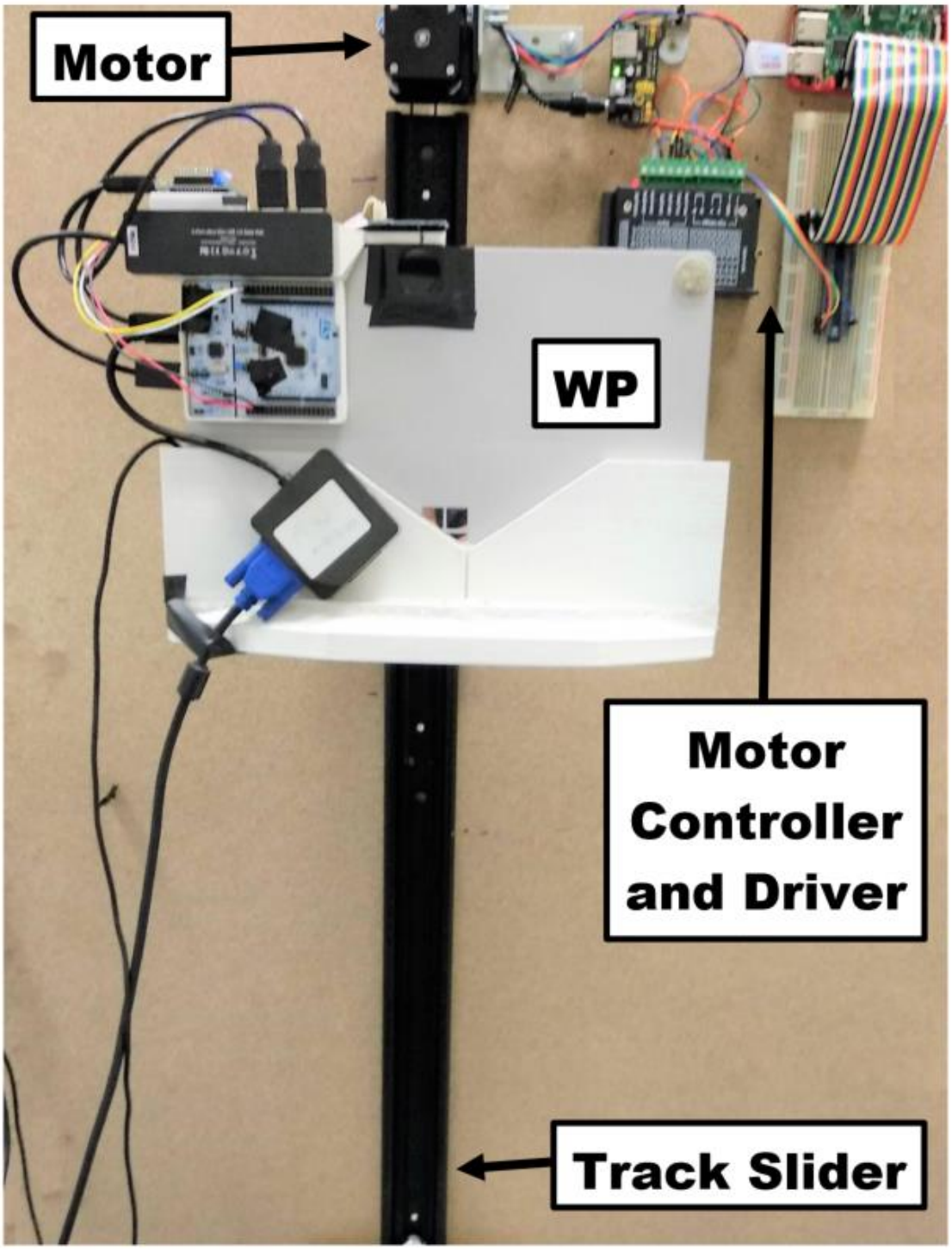


Figure 44: Experimental Setup – Mobile Case: WP mounted on Vertical Motorised Track Slider

4.4.5 *Mobile Case – Results*

The results of the experimental validation in mobile case are shown in Table 14. These results correspond to the results of simulations in scenario 3, shown Table 11. Likewise, a visual representation of the results of this experiment is shown in Figure 45, which corresponds to results of simulated scenario 3 shown in Figure 41.

These results bear a strong resemblance to those of the corresponding simulations. The RMSE in P_z^W was the highest of the three position elements of the WP. Also, it was higher than that in P_y^W by a comparable ratio of approximately 50 %. Likewise, the RMSE in P_x^W had the lowest value of the three position elements of the pose vector P_{WP}^W. Overall, the RMSE was lower than that in the corresponding simulated scenario 3. The discrepancy between these results was low and in the order of several millimetres, i.e. less than 5 mm. One of the reasons for such as low value of RMSE is the relatively low velocity of the WP whose period was $T = 34\ s$. Also, the motor controller ensured a smooth change of the motion's direction at the inflection points, i.e. when $y^W = -0.5$ or $y^W = 0.3$. It may have, to some extent, reduced the error in IMU readings. Moreover, this motion pattern involved no rotations, thus making the IMU readings less susceptible to error.

TABLE 14: MOBILE CASE - EXPERIMENTAL RESULTS

	$P_x^W [m]$	$P_y^W [m]$	$P_z^W [m]$	**Total [m]**
RMSE	0.0025	0.0115	0.0204	0.0136

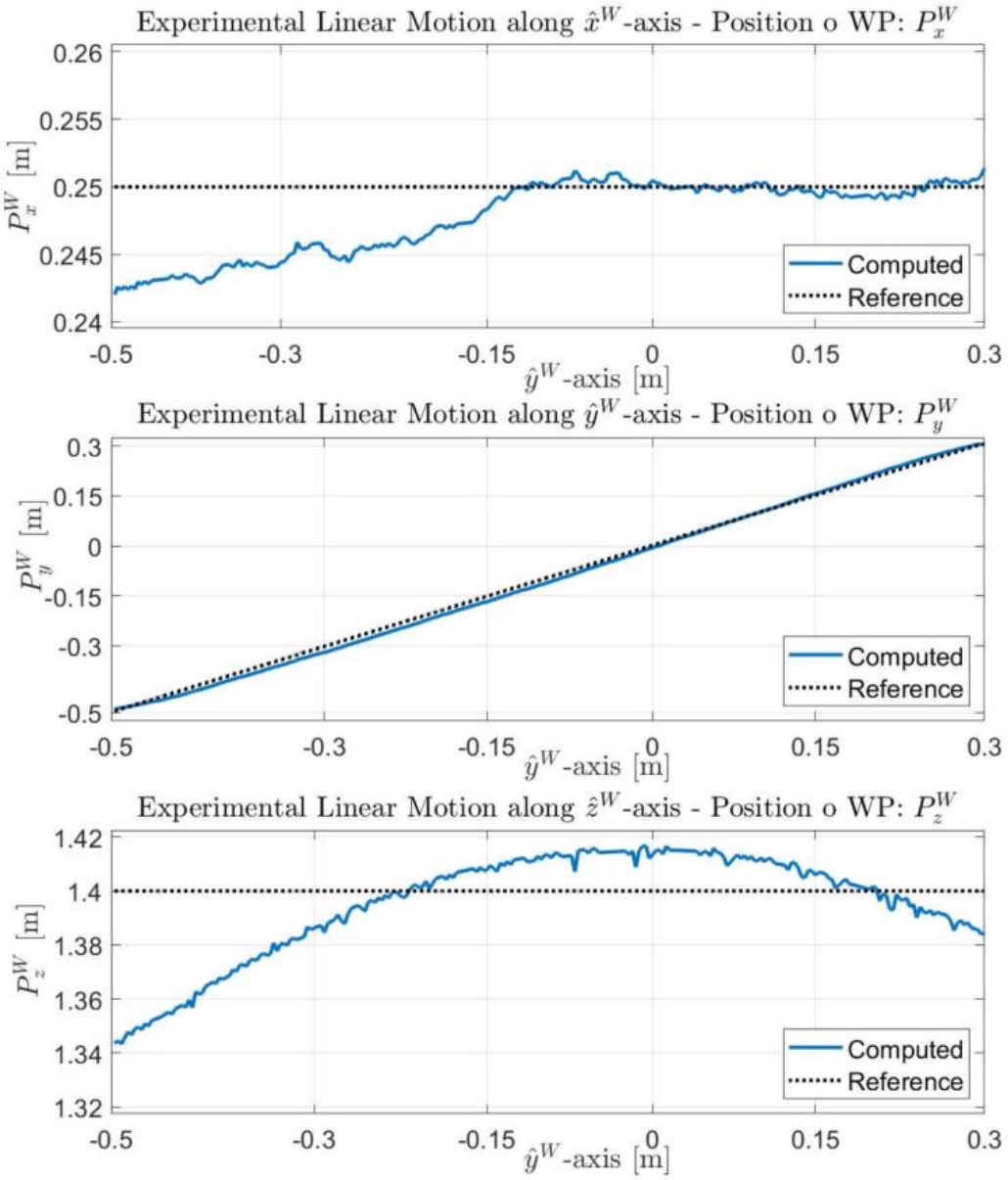


Figure 45: Experimentally Determined Position of the WP in Linear Motion along $\hat{y}^W$-axis

An additional experiment was carried out to determine the repeatability of the proposed system and its algorithm. To this end, the motor controller program on the Raspberry Pi was programmed to drive the WP to perform ten full cycles of scenario 3; to simulate ten repetitions of the barbell squat, which generally resembles a mostly straight, vertical, path, depending on where the WP is positioned. Figure 46 shows the results of this experiment. These results show that the performance of the proposed system was consistent and repeatable in all ten cycles. It is also evident that the output of the IMU did not drift, thus avoiding an adverse impact on the sensor fusion algorithm's accuracy.

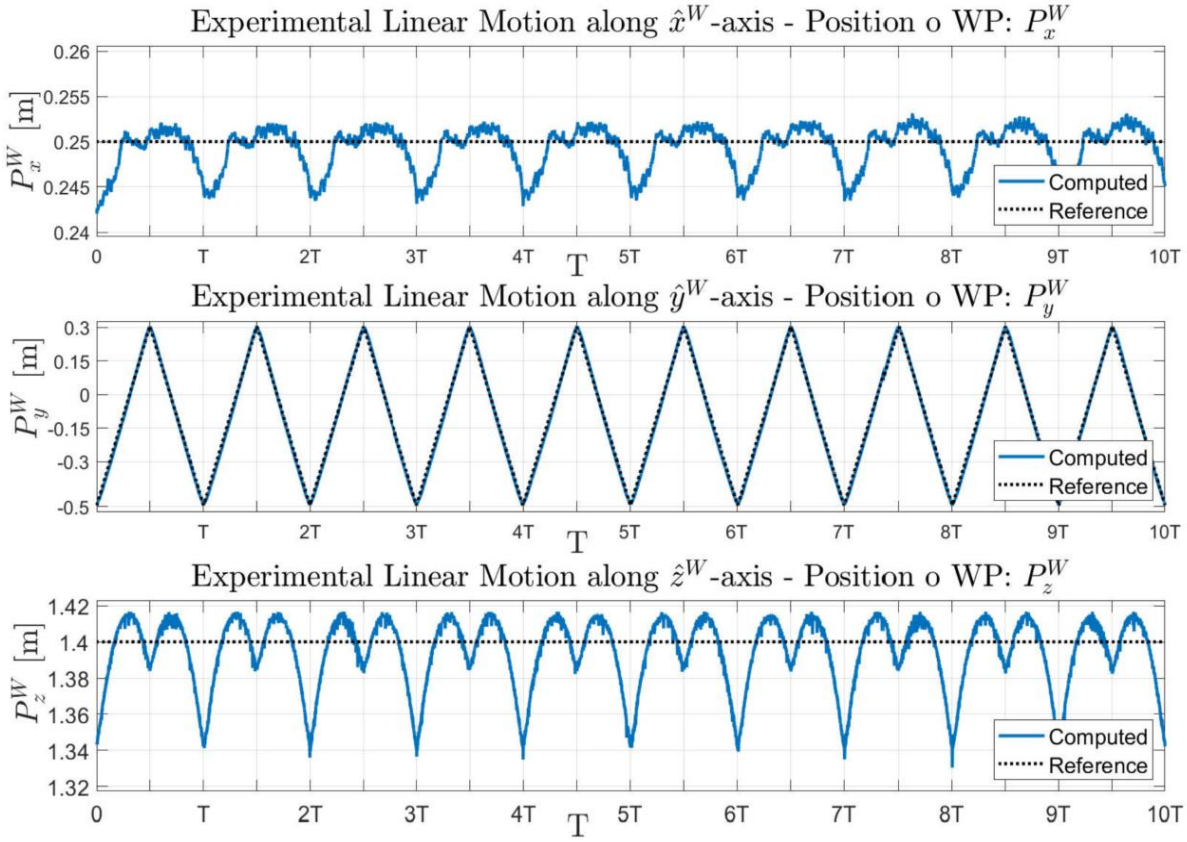


Figure 46: Experimentally Determined Position of the WP in Linear Motion along $\hat{y}^W$-axis Over Ten Repetitions with T = 34 s

4.5 Discussion of Results and Comparison with SOA

The performance of the proposed system was compared to similar systems that exist in the SOA, i.e. the opto-inertial trackers that relied on as few points of reference as possible. One of the key selection criteria for this comparison was the similarity in terms of system architecture, in particular the use of monocular vision and IMU sensor fusion for pose estimation. A direct one-to-one comparison was not possible due to different performance validation metrics, target application spaces, system architectures, cost, and the algorithms used in these approaches. However, a general comparison can be made. Table 15 compares and contrasts some of the key properties of the proposed system to the three most comparable alternatives in the SOA, as reported in the respective referenced publications.

TABLE 15: COMPARISON OF THE PROPOSED SYSTEM TO ALTERNATIVE SOLUTIONS IN THE SOA

	Position Error [mm]	Markers Required	Tracking Type	Work Envelope Size (along z-axis) [m]	Overall System Complexity
IS-1500 (PRA algorithm with Fiducial Markers) [11]	2 (Typical)	At least 4 (Passive Fiducial)	Inside-Out	Variable	High
Maereg et al. [13]	0.21 (Static) (RMSE)	2 (Active)	Outside-in	0.045	Low
Li et al. [12]	48.3 to 275.4 (Static) (RMSE)	2 (Passive)	Outside-In	1.13 to 4.13	Low
Proposed System	36.7 (Static), 13.6 (Mobile) (RMSE)	2 (Active)	Inside-Out	0.5 to 1.5	Low

The key metric to be evaluated was the overall error in position estimation of the wearable/mobile device in 3-D space, as well as the comparison of parameters that describe the key requirements of the individual systems. The IS-1500 tracking system was the most accurate inside-out tracker, whose position error was, by far, the lowest. However, this tracker required at least four fiducial markers and high external computing power capabilities to achieve such results, thus being the most expensive and complex system in this comparison. The opto-inertial motion tracking system proposed by Maereg et al. reported very low RMSE. However, it achieved such an accuracy within the smallest work envelope of only several centimetres and only in the static case, at a single position, while the accuracy in mobile case was not assessed quantitatively. Nevertheless, it was a low-cost outside-in tracker. On the other hand, the system proposed by Li et al., which was also an outside-in tracker, had a similar performance to the system proposed in this work. It was also validated in a somewhat similar way. The RMSE was determined at a number of static positions along a straight line parallel to the $\hat{z}^W$ axis at distances between 1.13 and 4.13 metres. However, the proposed system achieved lower overall RMSE in position estimation in both static and mobile cases, as shown in Table 13 and Table 14, respectively. Both systems had a low complexity. However, the tracker proposed by Li et al. was an outside-in tracker, while our proposed system was and inside-out tracker.

The proposed system advances the SOA in the following ways. It combines the advantages of the comparable alternatives in the SOA. Firstly, it is an inside-out opto-inertial tracking system. The advantage of an inside-out tracker over the outside-in trackers is in that the size of the work

envelope can be scaled at little to no expense. The costliest component, both in terms of price and complexity, is the camera. The proposed system, like the IS-1500, has one monocular camera embedded in the WP, regardless of the size of the work envelope. The algorithm does not change, as long as two points of reference are in camera's FoV and their baseline B is known. Whereas the outside-in systems would require additional cameras to scale the work envelope, the proposed system would need only additional IR LEDs, whose complexity and cost implications are significantly lower. Secondly, the proposed system developed and described in this thesis is less complex in terms of the architecture and algorithm, as compared to the IS-1500. In this regard, the proposed novel tracking system is more comparable to the two outside-in alternatives that also rely on two tracking points of reference. In summary, the proposed system has the advantage of the inside-out systems while being less complex and, thus more suitable for low-cost and low-power, miniaturized, battery powered wearable motion tracking devices for various application spaces, such as the barbell squat in ST.

The main limitation of the proposed system is in that the highest accuracy is achieved when the WP is near the centre of the work envelope and rotates mainly about a single axis while the rotations about the remaining two axes are relatively small. Extreme poses in the WP increase the RMSE, specifically those with high rotation angles, which was shown in the static case of the experimental validation which is related to the trigonometric functions used in the sensor fusion algorithm used to compute the 3-D pose. Nevertheless, the proposed sensor fusion algorithm can handle multi-axis rotations with rotation angles up to approximately $10\ deg$ about each axis, which was shown in the simulated uniform random scenario 2. The proposed algorithm is susceptible to noise in IMU readings. The point noise also affects the performance but to a lesser degree. The impact of noise $\mathcal{N}$ is particularly high in scenarios that involve significant multi-axis rotations, such as that in the simulated scenario 2, whose impact is shown in Figure 39.

4.6 Conclusions and Summary

In this work, a system architecture for low-power, miniaturized, wearable human motion tracking systems for sports applications was presented. The proposed system comprised of the WP which incorporated two sensor modalities, i.e. a monocular camera and an IMU sensor. The WP used two points of reference embedded in the ambient environment, i.e. IR LEDs. Furthermore, a novel multimodal sensor fusion algorithm for the proposed system architecture was presented. The WP is an inside-out-tracker. The sensor fusion algorithm runs on the WP, which leverages

the complementary nature of the monocular vision and IMU sensor modalities to directly compute the 3-D pose of the WP. The target application spaces for this system include sports applications. It can be particularly applicable to tracking certain exercises in ST routines, such as the barbell squat which follows a prescribed series of movements.

This work proposes an alternative approach to human motion tracking using wearable devices. It proposes an inside-out opto-inertial motion tracker that performs 3-D pose detection using only two points of reference in the ambient environment. It is a less expensive, simpler, and more scalable approach, as compared to the alternatives present in the SOA, such as the IS-1500. On the other hand, the two outside-in trackers considered in this work are less scalable, while being similar conceptually. Also, their usability in the context of wearables is limited by the fact that their accuracy is also affected by the distance between the two points of reference, which must be small if these were to be attached to the human body. Moreover, the small distance between the reference points, in conjunction with considerable distance away from camera, increases the cost of the system, due the requirement of a higher camera resolution to maintain the precision of point detection. Thus, the proposed tracker advances the SOA by proposing a new alternative to the existing systems, albeit not as accurate as the leading IS-1500. However, it can be considered a viable alternative if other factors are taken into account, such as the cost or scalability which are important considerations in many application spaces such as the ST; considered in this work. Moreover, the proposed system achieved a sufficiently low error in position estimation to be good enough for tracking human motion in certain exercises, such as the barbell squats in ST routines.

The proposed system was implemented and validated in the form of a prototype experimental setup in laboratory conditions. Its performance was experimentally validated in two scenarios, static and mobile. The static case was aimed at determining the performance in terms of accuracy across the entire work envelope. The mobile case focused on the motion pattern that is normally involved in a barbell squat. This scenario was of the primary interest, as this system is intended to be used in tracking such motion patterns when it has moved to the next development stage, i.e. a small-form-factor prototype stage implementation giving real-time information about body posture and position.

The proposed system compared well to the other two outside-in tracking systems, as shown in Table 15. It needs to be noted, however, that the RMSE of these two systems cannot be directly compared due to different validation scenarios. Therefore, the experimental conditions need to

be also taken into account. Nevertheless, the proposed system performed better than that proposed by Li et al. Although the monocular version of their system was validated at a set of static positions along a straight horizontal line with no rotations, RMSE of the system proposed was lower in both experimental scenarios. On the other hand, the outside-in tracker proposed by Maereg et al. achieved lower RMSE. However, it achieved this result within a much smaller work envelope in static conditions with no rotations and is, thus, not directly comparable to the proposed tracker. The proposed system did not match the performance of the IS-1500 inside-out tracker, which had the lowest error of all comparable systems present in the SOA. The IS-1500 had the highest accuracy of all methods considered in this work. However, little detail is known about the methods used in evaluating this system.

The analysis of the processing speed was not described in this chapter for two reasons. Firstly, all work that was described in this chapter was carried out offline on high-powered PC-grade computers. The execution time of the proposed algorithm would not be representative of its potential in the context of low-power embedded systems. The real-time performance was evaluated using the embedded version of the prototype system, which is described in detail in the next chapter, i.e. Chapter 5.

5 Embedded Prototype Multimodal Tracking System

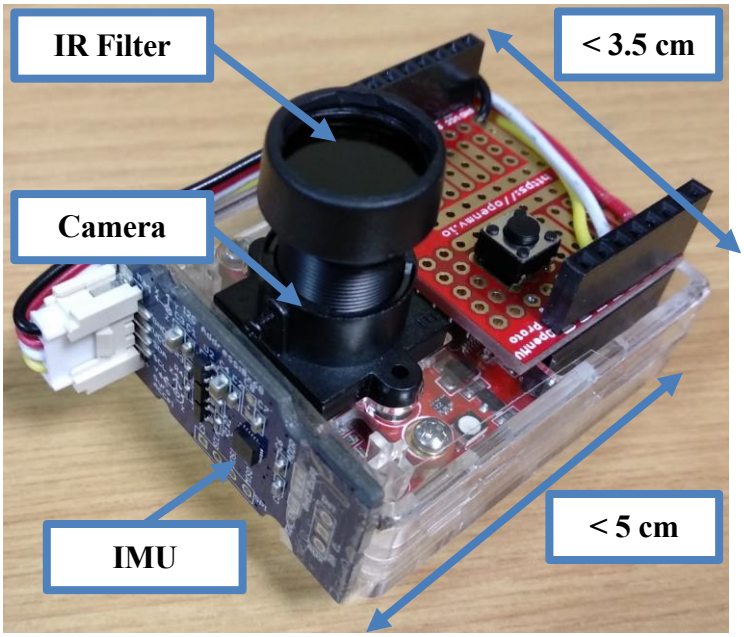
M. P. Wilk, M. Walsh, and B. O'Flynn, "Low Cost Embedded Multimodal Opto-Inertial Human Motion Tracking System", 31st Irish Signals and Systems Conference (ISSC), 2020, (accepted)

5.1 Multimodal Tracking

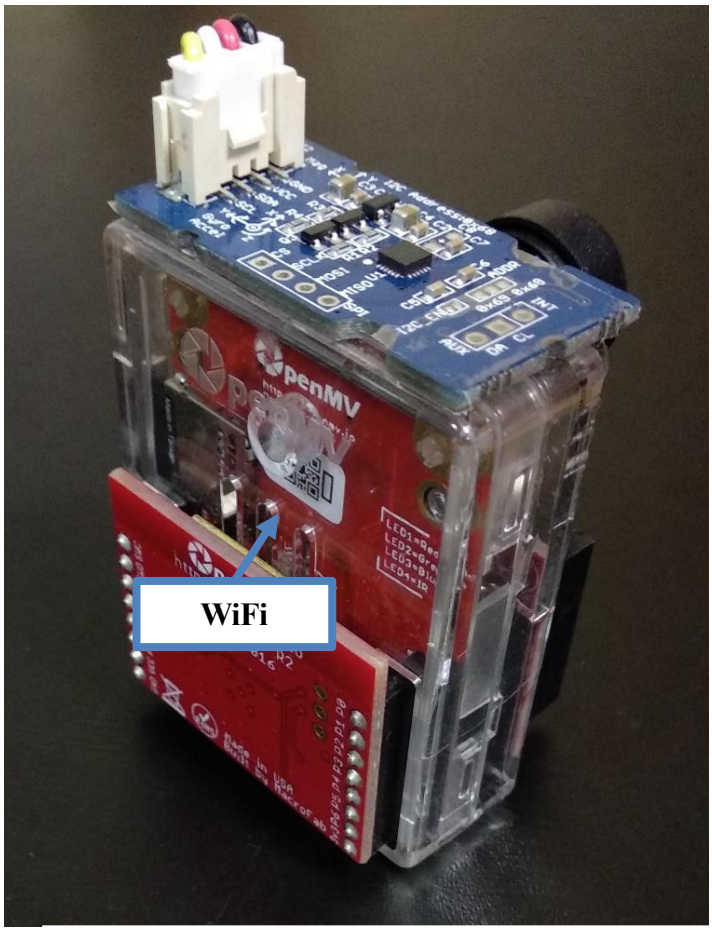
A demonstrator of the proposed system was developed as an embedded proof-of-concept prototype to test the main hypothesis of this work, as described in section 2.2. As outlined in the hypothesis, the demonstrator system is a low-cost embedded system that incorporates vision and IMU technologies in an MCU based wearable device, i.e. the WP, that runs the proposed novel algorithms for 3-D motion tracking with two external reference points placed in the ambient environment. The 3-D pose is computed by fusing two information from two sensor modalities that complement each other, i.e. camera and an IMU. The objective was to test if it was feasible to successfully implement the proposed system in the context of low-cost and resource-constrained conditions, thus proving the hypothesis of this thesis true. One of the possible application spaces includes tracking motion in ST exercise routines. A specific example of an ST exercise includes a barbell squat which the proposed system can be for tracking. The barbell squat involved a repetitive motion pattern that is generally a slight curve in 3-D space along a vertical axis. The WP can be attached to the back of the athlete such that its camera faces the IR LEDs. The 3-D pose computed by the WP (or multiple units of WP attached to different parts of the back) can be used as an input to another system that could use aspects of machine learning to determine how closely the measured motion pattern matches an expected (correct) motion patterns. However, the scope of this research work was limited to the development and validation of the described motion tracking system.

The prototype implements the proposed system architecture and the novel multimodal sensor fusion algorithm for 3-D pose detection including point detection. A small form-factor platform was selected for wearable applications. The OpenMV Cam H7 development board was used in this task [125]. This is an MCU-based platform designed for rapid prototyping of projects that incorporate machine vision. It uses an Arm Cortex-M7 STM32H743VI MCU, which is powerful enough to perform real-time image processing tasks while being embeddable in a small, low-cost, and energy efficient wearable device, referred to as WP [126]. The WP system included two

sensor modalities, i.e. vision and IMU. To this end, the prototype incorporated the MT9V034 global shutter camera module which comprises of the image sensor and a fixed-focus lens [49]. The camera module includes a custom developed optical IR filter which was attached to the camera's lens, to allow the camera to detect the IR light spectrum only; which is at the same wavelength as that emitted by the IR LEDs emitted [94]. The IMU used for the implementation was the same as that used in the experimental validation described in Section 4, i.e. the MPU9250 [48]. Additionally, a WiFi shield was added to enable wireless communications [127]. The miniaturised prototype is shown in Figure 47 with the associated building blocks of the proposed technology. The complete demonstrator of the location tracking system is shown in Figure 48. Although it is not a strictly wearable system in this form, it can be considered as one. The small form factor of the WP and its ability to operate wirelessly using a battery makes it a readily wearable device. This includes the WP as well as the two points of reference, i.e. the IR LEDs. The system uses these technologies and the novel algorithms developed as part of this work to locate and dynamically track the position and orientation of the WP, as described in Chapters 3, **Error! Reference source not found.**, and 4. Although the WP was implemented using a general-p urpose off-the-shelf prototyping platform, it is clear that it can be also implemented in a significantly smaller form factor, thus making it even more suitable for wearable applications.



(a)



(b)

Figure 47: Demonstrator Prototype System

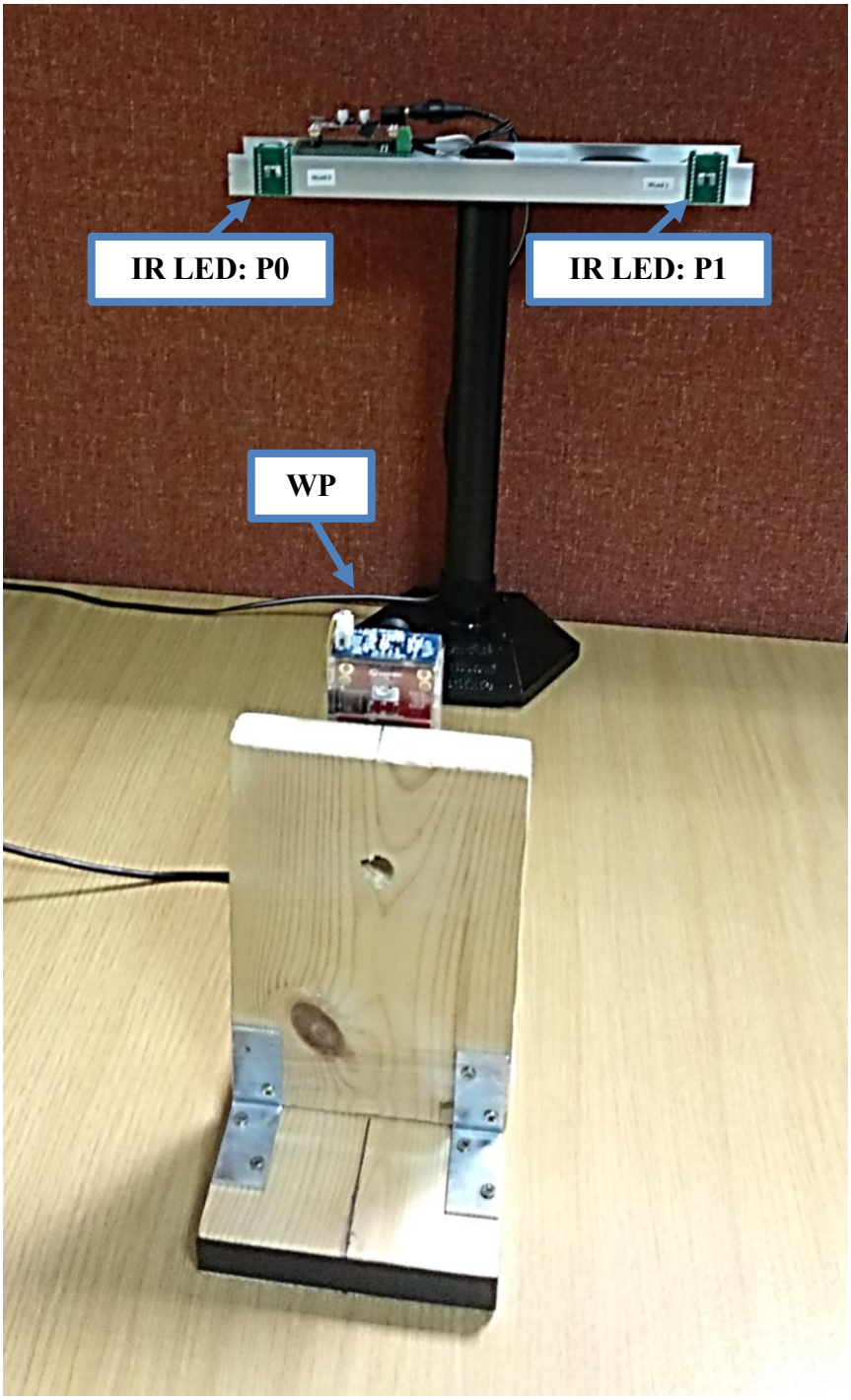


Figure 48: Demonstrator System

The WP can run embedded code written in MicroPython [128]. MicroPython is an implementation of the Python 3 programming language that contains a small subset of standard Python libraries optimised for resource-constrained MCUs. The proposed novel algorithm, described in detail in Section 4, was implemented in MicroPython as a custom class that could

be imported into the main program's file. Figure 49 shows a screenshot of the OpenMV IDE during one of the tests of the proposed system in the setup shown in Figure 48. The 3-D pose was computed based on the input vectors defining the IMU rotation angles and positions of the two points of reference on the image plane, $Thzyx$ and $P01$, respectively; as described in Section 4.2.2. For each image frame, the pose was computed by making a call to the member function of the pose3D object of the custom class which contained the implemented algorithm, as shown in the highlighted line of code in Figure 49. This line of code corresponds to the Data Fusion Block shown in Figure 50. This block diagram was described in detail in Section 4.2.1. The Data Fusion block contains the proposed multimodal sensor fusion algorithm while the other blocks describe the tasks that are carried out in order to condition the input data prior to passing it to the main Data Fusion Block. The input parameters $P01$ and $Thzyx$ correspond to p^I and θ^W, respectively; except for the order or elements in $Thzyx$ which were reversed. The order or elements in $Thzyx$ was reversed to maintain the consistency with conventions adopted in the embedded code, as compared to the theoretical derivations described in Section 4.2.2.

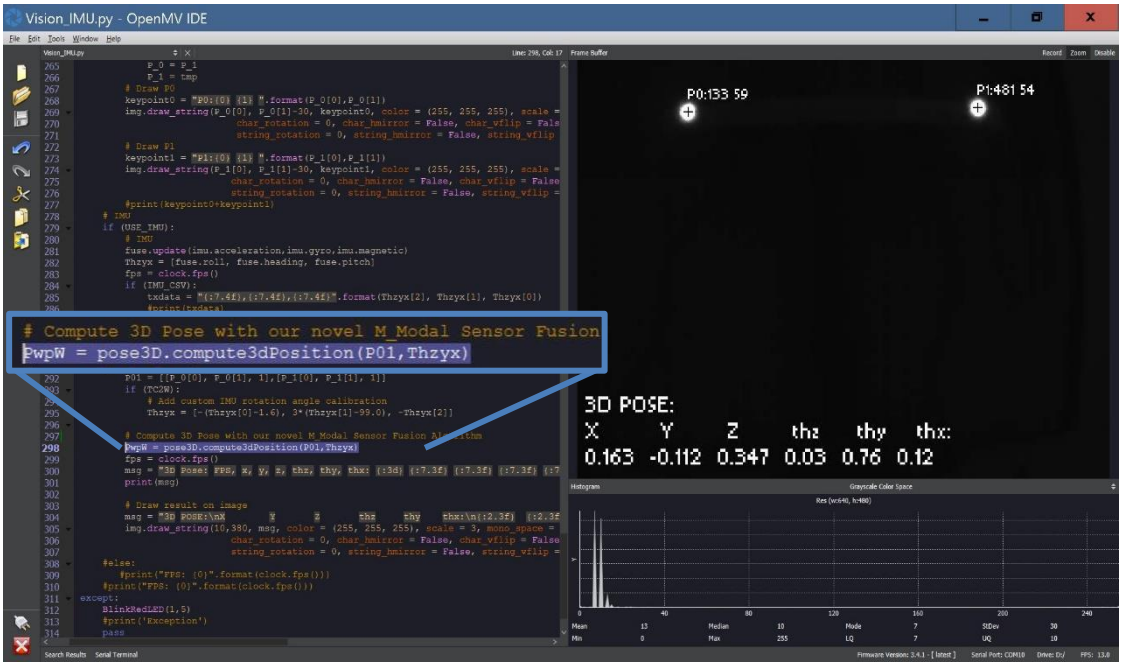


Figure 49: Embedded Execution of the Proposed Multimodal Sensor Fusion Algorithm in OpenMV IDE in Real-Time

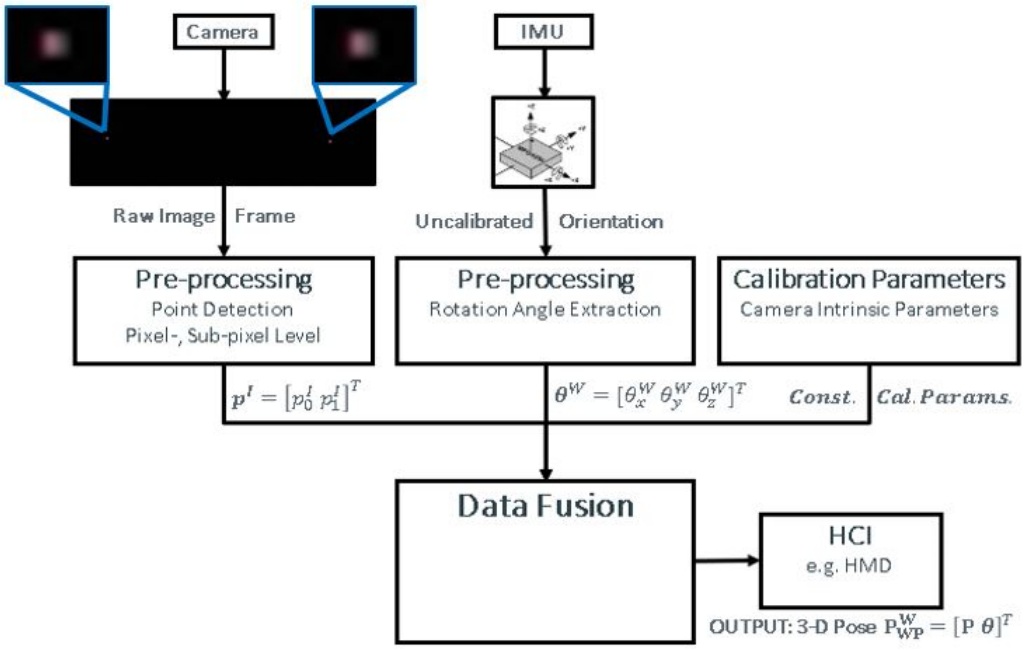


Figure 50: General Block Diagram of the Proposed Multimodal Sensor Fusion Algorithm, along with the Input Pre-Processing Stages

5.2 Performance Evaluation

An initial evaluation of the proposed system has been completed in terms of processing speed and accuracy. The objective of this process was to replicate the experimental work that was carried out on the larger, pre-prototype-stage, (non-wearable) version of the proposed system, and in particular the validation of the proposed multimodal sensor fusion algorithm, as described in Chapter 4, in a wearable miniaturised system. The performance in the mobile scenario was of particular interest, given the cosidered target applications space in sports and ST. the specific exercise to be simulated was the barbell squat in the context of ST.

5.2.1 *Experimental Setup*

The resource-constrained miniaturised wearable embedded system was evaluated in a mobile case scenario, i.e the Scenario 3 in Section 4.4.4. However, the conditions in this evaluation procedure had several differences which were casued by the use of different camera modules in the two versions of the experimental WP. Whereas the first version, described in Chapter 4, had the module OV8865 [92], the embedded version of the WP had the camera module MT9V034 [49]. Among many differences, the MT9V034 camera module had a smaller FoV which required

different positioning of the embedded WP with respect to the IR LEDs, as compared to the previous experiments with the non-wearable version of the WP, described in section 4.4.4. Firstly, the vertical slider was positioned farther away from the IR LEDs and the range of motion along the $\hat{y}^W$ was slightly different. Specifically, the WP was positioned at $z^W = 1.5\ m$ and the $y^W \in < -0.55, 015 >$ m, while $x^W = 0.25$ m, as compared to the settings of the previous experimental setup, i.e. $x^W = 0.25$ m, $y^W \in < -0.5, 0.3 >$ m, $z^W = 1.4$ m. The IMU was calibrated and the orientation angle measurements based on its output were used as an input in computing the 3-D pose in real-time. Likewise, the camera on the WP was set to resolution of 640-by-480 pixels, and calibrated using camera calibration [61]. Subsequently, the elements of the intrinsic parameter matrix were used to set the constants in the proposed algorithm by hardcoding them in the MicroPython implementation of this algorithm which are also shown in Figure 50. However, the input images were no corrected for lens distortion. The number of 3-D pose measurements in the experiment was $N > 4000$, while the framerate was more than 22 FPS. The main differentiator in this process, as compared to the previous experiments described in section 4.4, was the fact that the WP computed the 3-D pose in real-time. The WP performed all the functions described in the block diagram in Figure 50 in real-time during the data acquisition procedure, while the slider was in motion. The WP that was mounted on the vertical slider in this experimental setup is shown in Figure 51 and described in detail in section 4.4.4.

Additionally, the execution time was determined to evaluate the processing requirements of the data acquisition and fusion. To that end, the IMU, vision and 3-D pose tasks were separated.

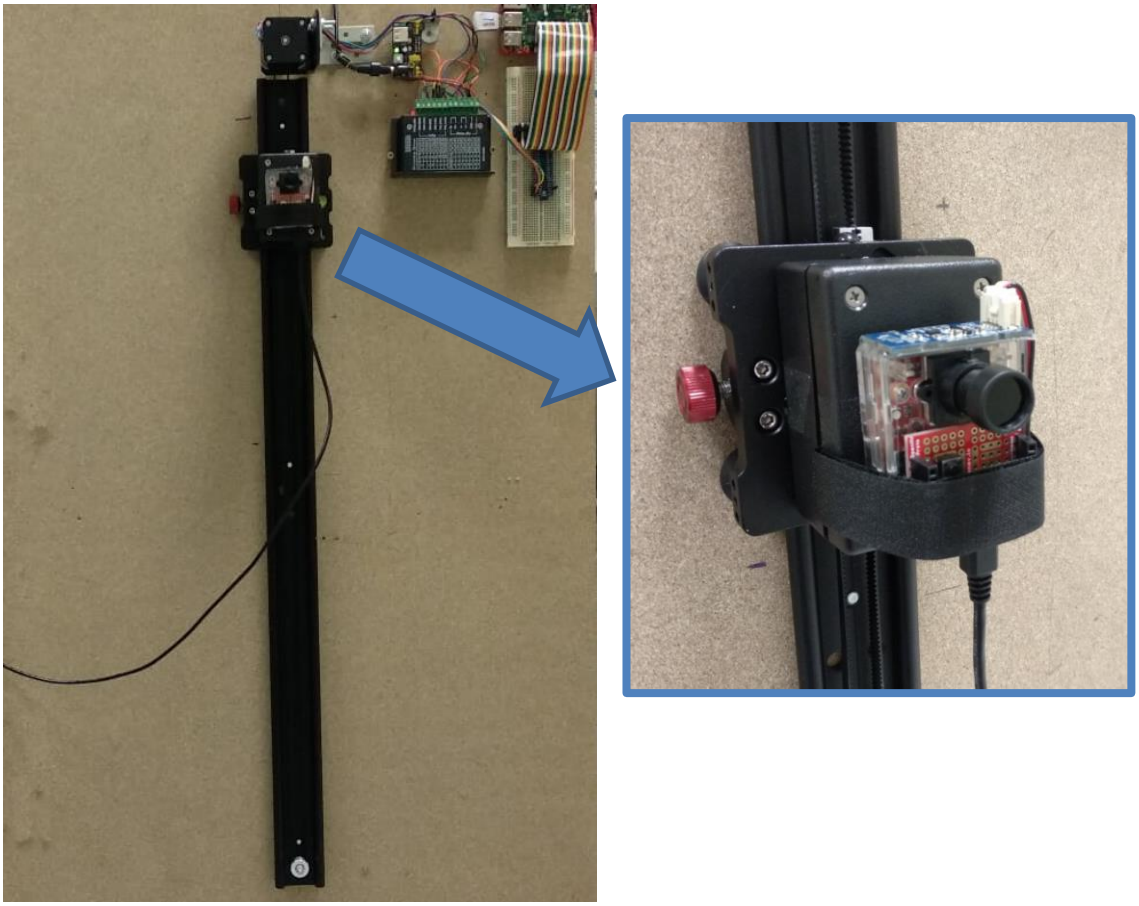


Figure 51: Experimental Setup: Small-Form-Factor WP was Attached to the Vertical Slider

5.2.2 *Results and Discussion*

Despite the resource-constrained nature of the platform, validation trials have shown that the performance, in terms of accuracy, of the embedded version of the system was generally consistent with that of the large-form-factor unit described in the previous chapter in Section 4.4. The acquired data points were processed following the steps from section 4.4.5. The key metric under scrutiny was the RMSE in position and orientation calculations, i.e. the 3-D Pose.

Figure 52 shows the output of the proposed embedded data fusion algorithm as a function of time, i.e. the position elements of the pose vector P^W over eight up-down cycles of the WP. It shows the accuracy in computing the 3-D position as well as the repeatability over time.

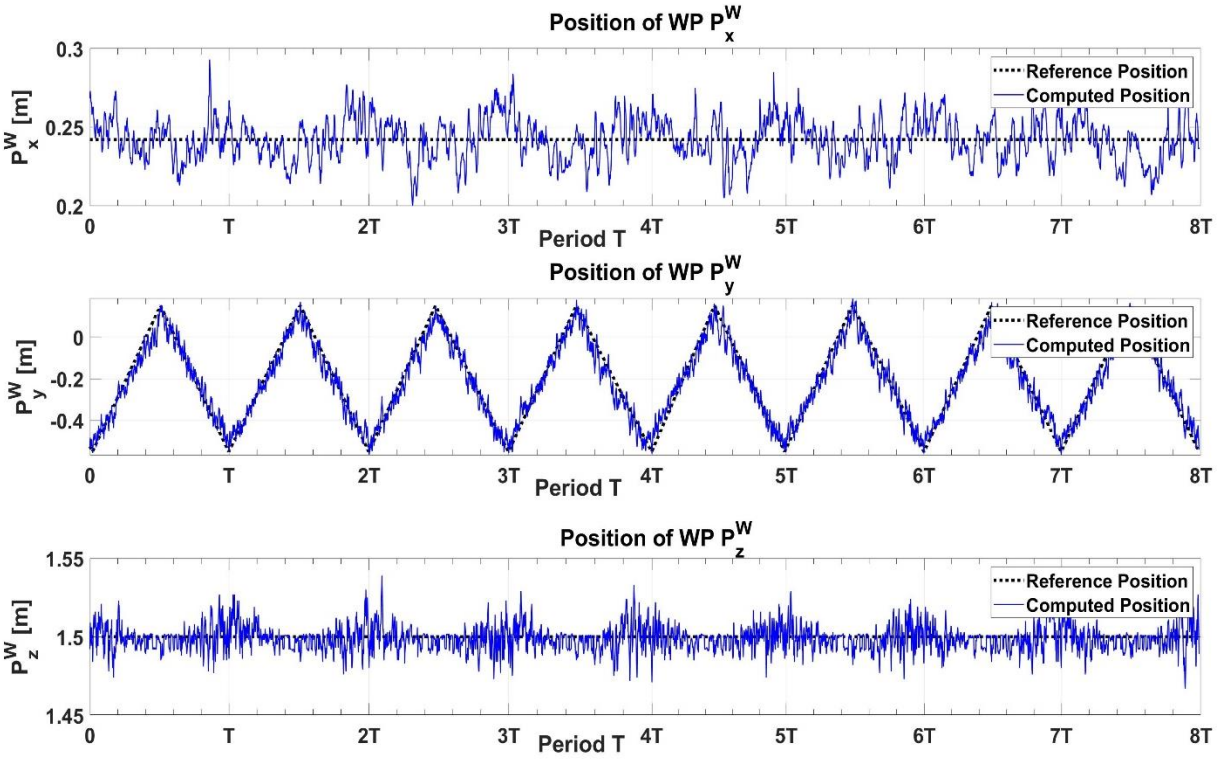


Figure 52: Experimentally Determined Position of the Small-Form-Factor Version of WP in Linear Motion along $\hat{y}^W - axis$

In these tests, the RMSE in the position and orientation calculations of the WP was computed and is shown in Table 16 and Table 17, respectively. Table 16 shows that the overall RMSE in position and orientation was 3.28 cm and 0.8921 degree, respectively.

The RMSE in orientation was determined in this experiment the algorithm proposed by Madgwick et al. [46]. The RMSE was generally within the expected range, except for θ_y^W; which was 1.4322 degree.

The total RMSE in position was greater than the corresponding results from the previous chapter shown in Table 14 which was measured at 1.36 cm. Considering the differences between the two experimental platforms this is to be expected. Firstly, the miniaturised wearable version of the WP used a different camera with a much smaller resolution, i.e. 640x480 pixels, as compared to 3264x2448 used in the large-form-factor version of the WP. Secondly, the WP was positioned at $z^W = 1.5\,m$, which was 10 cm farther away from the origin of the World frame of reference. Thirdly, this version of the WP computed the 3-D pose in real-time using the resource-constrained embedded system. Therefore, it was necessary to facilitate the maximum framerate to capture the motion in this mobile scenario. To this end, the input images were not corrected for lens

distortion; to reduce the computational requirements associated with processing the input frames. This fact, along with the relatively high RMSE in the rotation angle about $\hat{x}^W$-axis θ_x^W, shown in Table 17, contributed to high RMSE in P_y^W. As a result, the RMSE in P_y^W was high which significantly increased the total RMSE in position computation. The remaining two position elements of the 3-D pose vector P^W were low, with P_x^W being higher while P_z^W lower than their corresponding values in the previous chapter, shown in Table 14, 0.25 cm and 2.04 cm, respectively. In fact, the RMSE in P_z^W was not expected to be lower in this experiment, given the fact that WP was positioned 10 cm farther away from the origin of world reference frame.

TABLE 16: EXPERIMENTAL VALIDATION: RMSE IN POSITION CALCULATION

	$P_x^W[m]$	$P_y^W[m]$	$P_z^W[m]$	**Total [m]**
RMSE	0.0136	0.0545	0.0080	0.0328

TABLE 17: EXPERIMENTAL VALIDATION: RMSE IN ORIENTATION CALCULATION

	$\theta_x^W[deg]$	$\theta_y^W[deg]$	$\theta_z^W[deg]$	**Total [deg]**
RMSE	1.4322	0.4444	0.3719	0.8921

One of the additional findings of this work was that the framerate was limited mainly by the image processing tasks. The execution time breakdown is shown in Table 18. The point detection was the most time consuming task. It is understandable since the algorithm needs to look for the points in all pixels in the image. A further reduction of the camera resolution and/or addition of the temporal element to point tracking algorithm to look for the points only in the areas of the image where they are expected to be, based on previous frames, could significantly decrease the execution time of this task.

TABLE 18: EXECUTION TIME BREAKDOWN

Task	**Execution Time [ms]**
IMU: Orientation Computation	1.3
Vision: Point Detection	30.3
Sensor Fusion: 3-D Pose Detection	17.5

5.3 Conclusions Regarding the Wearable Miniaturised Data Capture System

An embedded, wearable, miniaturised, low cost, proof-of-concept, version of the motion tracking system proposed was developed for experimental testing. It used the MCU based machine vision development board OpenMV H7 and the IMU MPU9250. Validation trials have been carried out to measure the accuracy of this wearable solution. The experiments simulated a mobile scenario in the considered sports application space using the example of a barbell squat in an ST routine. The use case in which the WP is attached to an athlete's back while executing the barbell squat was simulated to track its motion in a repetitive and controlled manner. The accuracy of the embedded version of the WP was consistent with the expectations which were based on simulations and experimental work that is descri bed in Sections 4.3 and 4.4 in Chapter 4. The overall RMSE of the embedded system was 3.28 cm. Despite using different hardware components and configuration of the embedded WP, the proposed novel sensor fusion algorithm computed the 3-D pose with a comparable accuracy to that achieved in the corresponding experiment described in Section 4.4. It remained in single-centimetre range of RMSE which can be sufficient for low-cost tracking of certain ST exercises, such as the barbell squat as described. Some of the key metrics of a proper squat technique include the squat depth, trunk forward lean or lateral hip shift whose range of motion is significantly higher and are up to 1 m, 35 cm or 30 cm, respectively; as described in detail in section 4.3.3. Moreover, the embedded WP performed the motion tracking function in real-time at over 20 fps. The system performed all tasks within 50 ms for each output 3-D pose which incluided image processing, IMU orientation, and the novel multimodal sensor fusion. The image processing involved in point tracking consumed the most amount of time. The proposed sensor fusion algorithm computed the 3-d pose within 17.5 ms. The achieved frame rate is sufficient for most ST exercises since most of them are executed relatively slowly, as compared to more dynamic sports disciplines where the a significantly higher frame rate is required. In practical terms, one repetition of a barbell squat has a duration of approximately 2 seconds, while it is almost impossible to complete one repetition of this exercise correctly in less than 1 second. Given that the proposed system achieved a frame rate of over 20 fps, it can make approximately eighty 3-D pose updates for each repetition of this squat which is sufficient for assessing the exercise form. The results also show that the motion tracking can be carried out effectively with a the described system. The prototype system described in this chapter is based on inexpensive general-purpose microcontroller chip that achieved real-time motion tracking with sufficiently high accuracy. In summary, the validation process, described in

this chapter, shows that the proposed system fulfils the performance expectations based on previous modelling and experimental work with the non-wearable prototype version, described in Chapter 4. Furthermore, the experimental results have shown that it has the potential to become a viable low-cost human motion tracking system for sports applications, such as strength training.

6 Thesis Summary and Conclusions

The objectives of work described in this thesis were twofold. Firstly, it was aimed at advancing the existing SOA by developing:

- A novel wearable opto-inertial human motion tracking system based on a multimodal sensor fusion and two external points of reference based on IR LEDs
- A novel multimodal sensor fusion algorithm for computationally efficient 3-D pose detection
- A novel subpixel point detection algorithm for lowering the processing requirement, for a motion tracking system by reducing camera's resolution while maintaining accuracy of point detection
- A novel reference point estimation algorithm for finding the locations of reference points used in validating subpixel point detection algorithms
- A novel proof-of-concept demonstrator prototype that implements the proposed system architecture and multimodal sensor fusion algorithm in a miniaturised wearable form factor

The proposed system advances the SOA by increasing the feasibility of using such motion trackers in ST applications and other applications with similar requirements. It could act as an affordable alternative to existing systems which are more complex such as those used in motion capture labs. Existing systems that track 3-D pose tend to be complicated and generally expensive. In terms, of the comparable inside-out opto-inertial trackers, the leading alternative, i.e. the IS-1500 [11], is not widely available, while a single Antilatency tracker unit with a 3x3 metres tracking area option costs $385 [76] at the time of writing this document.

One of the potential uses of the new system includes human motion tracking in ST exercises with defined motion patterns, such as the squat, which may contribute to helping tackle one of growing societal challenges of ageing population [129]. ST is a proven and recommended addition to regular physical activities for all people to live a longer and more importantly healthier life. With this in mind, the British National Health Service recommends that ST is carried out at least twice a week, on top of other exercises [130]. However, ST can lead to injuries if executed incorrectly. Therefore, a professional sports coach is required to guide the individuals. However, the

accessibility and affordability of coaches might be limited with the growing proportion of population being older.

The proposed system can help ease this problem because it can be implemented as a low-cost wearable motion tracking device that could be used by individuals involved in ST to track their motion in real time and help ensure a proper exercise technique for the given individual is maintained.

Chapter 2 describes a review of motion tracking technologies that exist in the SOA. It began with reviewing various unimodal systems. Subsequently, multimodal systems were reviewed. Finally, a gap in the SOA was identified with respect to multimodal wearable human motion tracking systems, and a hypothesis for this work, described in this thesis, was formulated. Additionally, potential uses for this technology were explored.

In Chapter 3, a novel subpixel point detection algorithm, SLI, was presented to reduce the processing requirements related to point detection tasks in image processing. It can be used to determine the coordinates of points of interest in input images at the subpixel level, thus overcoming the limitation of the camera's resolution. It can compute the peaks of points in images faster and more accurately than the existing alternatives in the SOA under specific conditions, i.e. the points of interest were IR LEDs with specific intensity and wavelength. The significance of SLI in the context of this work was in that its use could significantly reduce the requirements related to the optical tracking component of the WP. The SLI enabled a reduction in the resolution of the camera in the WP by a factor of at least 2 without sacrificing the precision of the point detection algorithm. It also translated into cost reduction, because low resolution cameras are less expensive. Moreover, the lower resolution of images meant that image processing tasks can be executed much faster, which helps increase the frame/update rate of the system's output; a critical parameter of any motion tracking system.

Chapter 4 presents a novel algorithm for estimating the locations of reference points in images that are required to validate the accuracy of subpixel point detection algorithms under laboratory conditions. A quantitative validation of a subpixel point detection algorithm requires a reference point against which the results of the given algorithm are validated. A reference point is also necessary for comparing the accuracy of various subpixel point detection algorithms. In summary, this novel algorithm provides the means for performing these validation procedures. The chapter describes how the proposed algorithm was formulated and evaluated.

Chapter 5 is central in this work. It presents the novel multimodal sensor fusion algorithm that can be used to efficiently compute the 3-D pose of the WP in space. The proposed algorithm leveraged a system architecture that was based on a low-cost, miniaturised, wearable, opto-inertial motion tracking unit, i.e. the WP that incorporated two sensor modalities, the monocular camera and the IMU. The camera in the WP was used to track two known external points of reference, i.e. two IR LEDs. The proposed sensor fusion algorithm used the unique geometry formed by the camera and the reference points and complemented the missing information with rotation angles obtained from IMU's data. Thus, the 3-D pose of the WP in space was computed. Due to the specific system architecture, the mathematical formulations involved in pose detection were significantly simplified. The proposed system was evaluated in a series of simulated scenarios as well as laboratory conditions using a pre-prototype-stage platform that implemented its key functionalities. This chapter describes all aspects the proposed algorithm, system architecture, and its validation process.

Chapter 6 is a short chapter that focuses on the final development of a miniaturised proof-of-concept prototype system. The prototype system was based on works described in Chapter 5. It was intended to be used as a demonstrator of the complete proposed system in a wearable small form factor. It was also aimed at testing the hypothesis of this work and proving that such a solution can indeed close the identified gap in the SOA, thus significantly advancing it. Furthermore, it serves as the basis for future works that would focus on further development and validation of the proposed system.

In conclusion, the hypothesis of this work was tested and proven correct. The hypothesis is as follows: *We consider a low-cost, low-resolution, monocular camera system that is combined with an IMU in a single miniaturised wearable smart sensor unit, and it was coupled with two stationary points of reference, using active markers such as IR LED. Then the 3-D pose, i.e. the 3-D position and orientation, of the wearable unit could be efficiently determined. This approach has not been reported in existing literature. Moreover, the orientation data from the IMU could be used to directly complement the missing pieces of information from the vision sensor, thus reducing the overall system complexity; by avoiding the need for computationally expensive algorithms for computing the 3-D pose, such as the PnP solutions. As a result, the complexity of the sensor fusion algorithm for the 3-D pose estimation can be reduced and, thus, lead to lower requirements in terms of processing power and energy consumption. These requirements can be further decreased by reducing the computational load associated with the image processing tasks*

when detecting points of reference in images acquired by the camera. To that end, resolution of the camera can be reduced while introducing subpixel point detection techniques to finding the coordinates of the two points in the input images. The subpixel point detection can prevent the loss of precision of point detection caused by lowering camera's resolution. This results in a less complex and less expensive inside-out motion tracking system, as compared to the IS-1500 tracker. The 3-D motion can indeed be tracked efficiently by integrating a low-cost monocular camera and an IMU in a wearable opto-inertial tracker in the context of multimodal sensor fusion. The camera in the wearable tracker, the WP, can be used to track two external points of reference. The sensor fusion algorithm can use the geometry formed between the two points of reference and the camera and complement the missing pieces of information with IMU readings; to perform the 3-D motion tracking with a reduced computational complexity, thus leading to a lower cost of the system. Moreover, the proposed subpixel point detection algorithm contributes to further reduction of the cost of the system as the resolution of the camera can be reduced while maintaining the precision of point detection algorithm. This in turn has a direct impact on the processing requirements of the WP allowing a low-cost architecture to be used. Thus, the proposed system offers a viable alternative to the more expensive alternatives in the SOA. Although it did not achieve higher accuracy in terms of positional tracking than the leading systems in the SOA, it is accurate enough for many application spaces where affordability is an important consideration, such as the ST.

6.1 Key Contributions and advancements in the State-Of-The Art

The key contributions to the SOA are listed below, in the order of importance:

- A new multimodal sensor fusion algorithm for 3-D pose detection using wearable opto-inertial tracker and two external points of reference
- A new system architecture for efficient 3-D pose detection for human motion tracking applications
- A new subpixel point detection algorithm for efficient point detection at subpixel level to allow to for reduction of camera's resolution, thus allowing a user to use lower resolution of the camera without sacrificing the precision of point detection
- A new reference point estimation algorithm for finding positions of reference points used in future research activities

- A proof-of-concept novel demonstrator prototype that implements the proposed system architecture and multimodal sensor fusion algorithm

6.2 Future Work

The work described in this thesis offers opportunities to progress the field of research associated with low-cost multimodal sensing systems. This thesis was focused on developing the novel algorithms and proving the concept, as well as the accuracy and speed of the proposed system, which was achieved. However, there is room for improvement which is mainly an engineering task at this stage. Future work will involve further development and testing of the proposed system. Specifically, there are several key directions for these activities:

- **Demonstrator System:** The proposed system should be optimised. This involves mainly engineering tasks. One of the main tasks includes the optimisation of the implementation of the sensor fusion algorithm to maximise the framerate. This will include the implementation of event-driven or multithreaded software architectures to let algorithm run smoothly. This is necessary, because the current implementation, described in Chapter 6, assumes that the IMU and camera update rates are equal, which is generally not the case. Secondly, the image processing tasks need to be made computationally more efficient. To that end, the resolution of the camera needs to be reduced by a factor of 2 or possibly more. The loss of precision in locating the centres of points of interest can be prevented with the use of the proposed subpixel point detection algorithm. These improvements will at least double the framerate, from the current 20 FPS to more than 40 FPS. Also, a computationally efficient lens correction should be applied in future iterations to increase the accuracy of point detection. Perhaps, the lens correction could be applied to the two detected points as opposed to correcting the entire input frames which could significantly reduce the computational load associated with it. Finally, a more precise IMU and/or the accompanying orientation estimation algorithm could be considered, as error in orientation has an impact on the accuracy of the novel sensor fusion algorithm. At that stage, the demonstrator system will be ready for further validation and field trials.
- **Performance Validation with Human Subjects:** While the algorithms are proven and validated in lab testing, the proposed system and algorithms for position tracking described in this thesis need to be validated with human subjects in a “Gold Standard”

motion capture. Once its performance has been optimised, i.e. its framerate is high enough for performing human motion tracking in real-time, the system validation with human subjects can be carried out. Initially, an experimental protocol needs to be designed collecting data from individuals. The ST exercise, barbell squat, will be the first motion pattern to be used in validating the performance of the proposed system; with a motion pattern similar to that described in Chapter 5. Further experimentation is also recommended, especially such that can determine the performance in different use cases.

- **Miniaturisation and Embedded System Design:** Once the validation process has been completed, a highly miniaturised prototype system needs to be developed to make the vision of this project become reality. This work again will include mainly the engineering tasks. An embedded system could be developed that will comprise the miniature version of the WP and the two external points of reference, i.e. IR LEDs, as well as the novel algorithms described in this thesis, which are at the heart if this system.
- **Further Testing and Potential Commercialisation:** The miniature version of the proposed system needs to be further tested to prove its performance and potential commercial viability. The final activity of the future work is to explore the routes for potential commercialisation of this system. In this context, the IDFs (Invention Disclosure Forms) have been developed and submitted to the Technology Transfer Office in University College Cork, capturing the novel intellectual property developed and described in this thesis for potential exploitation through start-up or licensing activities.

References

[1] D. Lahat, T. Adali, and C. Jutten, "Multimodal data fusion: an overview of methods, challenges, and prospects," *Proceedings of the IEEE,* vol. 103, no. 9, pp. 1449-1477, 2015.

[2] R. A. Seguin, J. N. Epping, D. Buchner, R. Bloch, and M. E. Nelson, "Growing stronger; strength training for older adults," 2002.

[3] X. Guo, J. Liu, and Y. Chen, "FitCoach: Virtual fitness coach empowered by wearable mobile devices," in *IEEE INFOCOM 2017-IEEE Conference on Computer Communications*, 2017: IEEE, pp. 1-9.

[4] E. Muybridge, *The human figure in motion*. London: Chapman & Hall, LD, 1907.

[5] I. E. Sutherland, "A head-mounted three dimensional display," in *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, 1968: ACM, pp. 757-764.

[6] E. Foxlin, "Chapter 7: Motion Tracking Requirements and Technologies," *Handbook of Virtual Environment Technology, InterSense Inc., downloaded from prior to Feb,* vol. 28, p. 54, 2007.

[7] M. Windolf, N. Götzen, and M. Morlock, "Systematic accuracy and precision analysis of video motion capturing systems—exemplified on the Vicon-460 system," *Journal of biomechanics,* vol. 41, no. 12, pp. 2776-2780, 2008.

[8] A. Pfister, A. M. West, S. Bronner, and J. A. Noah, "Comparative abilities of Microsoft Kinect and Vicon 3D motion capture for gait analysis," *Journal of Medical Engineering & Technology,* vol. 38, no. 5, pp. 274-280, Jul 2014, doi: 10.3109/03091902.2014.909540.

[9] F. Schlagenhauf, S. Sreeram, and W. Singhose, "Comparison of Kinect and Vicon Motion Capture of Upper-Body Joint Angle Tracking," in *2018 IEEE 14th International Conference on Control and Automation (ICCA)*, 2018: IEEE, pp. 674-679.

[10] A. Filippeschi, N. Schmitz, M. Miezal, G. Bleser, E. Ruffaldi, and D. Stricker, "Survey of motion tracking methods based on inertial sensors: A focus on upper limb human motion," *Sensors,* vol. 17, no. 6, p. 1257, 2017.

[11] Intersense. "IS-1500." http://www.intersense.com/pages/70/255 (accessed 25.03.2019, 2019).

[12] J. Li, J. A. Besada, A. M. Bernardos, P. Tarrío, and J. R. Casar, "A novel system for object pose estimation using fused vision and inertial data," *Information Fusion,* vol. 33, pp. 15-28, 2017/01/01/ 2017, doi: http://dx.doi.org/10.1016/j.inffus.2016.04.006.

[13] A. Maereg, E. Secco, T. Agidew, D. Reid, and A. Nagar, "A Low-Cost, Wearable Opto-Inertial 6-DOF Hand Pose Tracking System for VR," *Technologies,* vol. 5, no. 3, p. 49, 2017.

[14] S. J. Byrnes, A. Lenef, F. Aieta, and F. Capasso, "Designing large, high-efficiency, high-numerical-aperture, transmissive meta-lenses for visible light," *Optics Express,* vol. 24, no. 5, pp. 5110-5124, Mar 2016, doi: 10.1364/oe.24.005110.

[15] E. L. Erickson, M. D. Kellam, P. R. Gill, J. Tringali, and D. G. Stork, "Miniature lensless computational infrared imager," *Electronic Imaging,* vol. 2016, no. 12, pp. 1-4, // 2016, doi: 10.2352/ISSN.2470-1173.2016.12.IMSE-269.

[16] L. Abraham, A. Urru, M. P. Wilk, S. Tedesco, M. Walsh, and B. O. Flynn, "Point tracking with lensless smart sensors," in *2017 IEEE SENSORS*, Oct. 29 2017-Nov. 1 2017 2017, pp. 1-3, doi: 10.1109/ICSENS.2017.8234060.

[17] Y. Wu, F. Tang, and H. Li, "Image-based camera localization: an overview," *Visual Computing for Industry, Biomedicine, and Art,* journal article vol. 1, no. 1, p. 8, September 05 2018, doi: 10.1186/s42492-018-0008-z.

[18] G. Xiao-Shan, H. Xiao-Rong, T. Jianliang, and C. Hang-Fei, "Complete solution classification for the perspective-three-point problem," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 25, no. 8, pp. 930-943, 2003, doi: 10.1109/TPAMI.2003.1217599.
[19] S. E. Leis and D. Ristau, "System for determining the spatial position and angular orientation of an object," ed: Google Patents, 1998.
[20] R. Klette and G. Tee, "Understanding human motion: A historic review," in *Human motion*: Springer, 2008, pp. 1-22.
[21] R. B. Martin, "A genealogy of biomechanics," in *23rd Annual Conference of the American Society of Biomechanics*, 1999, vol. 23.
[22] M. H. Pope, "Giovanni Alfonso Borelli—the father of biomechanics," *Spine,* vol. 30, no. 20, pp. 2350-2355, 2005.
[23] G. A. Borelli, *Joh. Alphonsi Borelli neapolitani matheseos professoris De motu animalium pars prima*. apud Petrum vander Aa, 1710.
[24] D. J. Sturman, "A Brief History of Motion Capture for Computer Character Animation," in *www.siggraph.org*, ed, 1999.
[25] J. C. a. A. P. T. W. Calvert. (1982, November) Aspects of the kinematic simulation of human movement. *IEEE Computer Graphics and Applications*. 41-50.
[26] C. M. G. a. D. Maxwell, "Graphical marionette," *Proc. ACM SIGGRAPH/SIGART Workshop on Motion,* pp. 172-179, 1983.
[27] M. Shamsi, M. Mirzaei, and S. S. Khabiri, "Universal goniometer and electro-goniometer intra-examiner reliability in measuring the knee range of motion during active knee extension test in patients with chronic low back pain with short hamstring muscle," *BMC Sports Science, Medicine and Rehabilitation,* journal article vol. 11, no. 1, p. 4, March 22 2019, doi: 10.1186/s13102-019-0116-x.
[28] R. States and E. Pappas, "Precision and repeatability of the Optotrak 3020 motion measurement system," *Journal of medical engineering & technology,* vol. 30, no. 1, pp. 11-16, 2006.
[29] Vicon. "Vicon Motion Capture Systems." https://www.vicon.com/ (accessed 20.05.2019.
[30] OptiTrack. "Motion Capture Systems." https://optitrack.com/ (accessed 20.05.2019.
[31] M. Yaman and S. Kalkan, "Multimodal Stereo Vision Using Mutual Information with Adaptive Windowing," in *MVA*, 2013, pp. 335-338.
[32] A. A. Mohamed, J. Baba, J. Beyea, J. Landry, A. Sexton, and C. A. McGibbon, "Comparison of strain-gage and fiber-optic goniometry for measuring knee kinematics during activities of daily living and exercise," *Journal of biomechanical engineering,* vol. 134, no. 8, p. 084502, 2012.
[33] H. Zhou and H. Hu, "Human motion tracking for rehabilitation—A survey," *Biomedical Signal Processing and Control,* vol. 3, no. 1, pp. 1-18, 2008.
[34] A. Alarifi *et al.*, "Ultra wideband indoor positioning technologies: Analysis and recent advances," *Sensors,* vol. 16, no. 5, p. 707, 2016.
[35] F. Adib, Z. Kabelac, D. Katabi, and R. Miller, "WiTrack: motion tracking via radio reflections off the body," in *Proc. of NSDI*, 2014.
[36] A. Banerjee, D. Maas, M. Bocca, N. Patwari, and S. Kasera, "Violating privacy through walls by passive monitoring of radio windows," in *Proceedings of the 2014 ACM conference on Security and privacy in wireless & mobile networks*, 2014: ACM, pp. 69-80.
[37] P. T. Anderson, "Electromagnetic tracking system and method using a single-coil transmitter," ed: Google Patents, 2007.
[38] P. T. Anderson, "Electromagnetic tracking system and method using a three-coil wireless transmitter," ed: Google Patents, 2006.

[39] J. Hummel *et al.*, "Evaluation of a new electromagnetic tracking system using a standardized assessment protocol," *Physics in Medicine & Biology,* vol. 51, no. 10, p. N205, 2006.
[40] L. K. de Paula, J. L. Ackerman, F. d. A. R. Carvalho, L. Eidson, and L. H. S. Cevidanes, "Digital live-tracking 3-dimensional minisensors for recording head orientation during image acquisition," *American Journal of Orthodontics and Dentofacial Orthopedics,* vol. 141, no. 1, pp. 116-123, 2012.
[41] D. Roetenberg, P. Slycke, A. Ventevogel, and P. H. Veltink, "A portable magnetic position and orientation tracker," *Sensors and actuators A: Physical,* vol. 135, no. 2, pp. 426-432, 2007.
[42] H. van Heeren and P. Salomon, "MEMS: Recent Developments, Future Directions," *Wolfson School of Mechanical and Manufacturing Engineering, Loughborough University, Loughborough,* 2007.
[43] D. J. Bell, T. Lu, N. A. Fleck, and S. M. Spearing, "MEMS actuators and sensors: observations on their performance and selection for purpose," *Journal of Micromechanics and Microengineering,* vol. 15, no. 7, p. S153, 2005.
[44] I. H. López-Nava and A. Muñoz-Meléndez, "Wearable Inertial Sensors for Human Motion Analysis: A Review," *IEEE Sensors Journal,* vol. 16, no. 22, pp. 7821-7834, 2016, doi: 10.1109/JSEN.2016.2609392.
[45] T. InvenSense, "MPU-9250, Nine-Axis (Gyro+ Accelerometer+ Compass) MEMS MotionTracking™ Device," ed, 2014.
[46] S. O. H. Madgwick, A. J. L. Harrison, and R. Vaidyanathan, "Estimation of IMU and MARG orientation using a gradient descent algorithm," in *2011 IEEE International Conference on Rehabilitation Robotics*, June 29 2011-July 1 2011 2011, pp. 1-7, doi: 10.1109/ICORR.2011.5975346.
[47] R. Mahony, T. Hamel, and J. M. Pflimlin, "Nonlinear Complementary Filters on the Special Orthogonal Group," *IEEE Transactions on Automatic Control,* vol. 53, no. 5, pp. 1203-1218, 2008, doi: 10.1109/TAC.2008.923738.
[48] T. InvenSense. "MPU9250." https://www.invensense.com/products/motion-tracking/9-axis/mpu-9250/ (accessed 29.04.2019, 2019).
[49] OpenMV. "Global Shutter Camera Module." https://openmv.io/products/global-shutter-camera-module (accessed 6.05.2019, 2019).
[50] S. Moon, Y. Park, D. W. Ko, and I. H. Suh, "Multiple Kinect Sensor Fusion for Human Skeleton Tracking Using Kalman Filtering," *International Journal of Advanced Robotic Systems,* vol. 13, Apr 7 2016, Art no. 65, doi: 10.5772/62415.
[51] M. Khorasaninejad *et al.*, "Visible Wavelength Planar Metalenses Based on Titanium Dioxide," *IEEE Journal of Selected Topics in Quantum Electronics,* vol. 23, no. 3, pp. 1-16, 2017, doi: 10.1109/JSTQE.2016.2616447.
[52] L. Abraham, A. Urru, N. Normani, M. Wilk, M. Walsh, and B. O'Flynn, "Hand tracking and gesture recognition using lensless smart sensors," *Sensors,* vol. 18, no. 9, p. 2834, 2018.
[53] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM,* vol. 24, no. 6, pp. 381-395, 1981.
[54] S. Q. Li, C. Xu, and M. Xie, "A Robust O(n) Solution to the Perspective-n-Point Problem," *Ieee Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, no. 7, pp. 1444-1450, Jul 2012, doi: 10.1109/tpami.2012.41.
[55] V. Lepetit, F. Moreno-Noguer, and P. Fua, "Epnp: An accurate o (n) solution to the pnp problem," *International journal of computer vision,* vol. 81, no. 2, p. 155, 2009.

[56] iFixIt. "Oculus Rift Constellation Teardown." https://www.ifixit.com/Teardown/Oculus+Rift+Constellation+Teardown/61128 (accessed 7.05.2019, 2019).
[57] O. Kreylos, "Oliver kreylos' research and development homepage-wiimote hacking," *Accessed Sep,* vol. 15, 2008.
[58] F. Ta. "6DOF Positional Tracking with the Wiimote." https://franklinta.com/2014/09/30/6dof-positional-tracking-with-the-wiimote/ (accessed 7.05.2019, 2019).
[59] K. Levenberg, "A method for the solution of certain non-linear problems in least squares," *Quarterly of applied mathematics,* vol. 2, no. 2, pp. 164-168, 1944.
[60] D. W. Marquardt, "An algorithm for least-squares estimation of nonlinear parameters," *Journal of the society for Industrial and Applied Mathematics,* vol. 11, no. 2, pp. 431-441, 1963.
[61] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on pattern analysis and machine intelligence,* vol. 22, 2000.
[62] M. Froehlich, S. Azhar, and M. Vanture, "An investigation of google tango® tablet for low cost 3D scanning," in *ISARC. Proceedings of the International Symposium on Automation and Robotics in Construction*, 2017, vol. 34: Vilnius Gediminas Technical University, Department of Construction Economics ….
[63] B. Khaleghi, A. Khamis, F. O. Karray, and S. N. Razavi, "Multisensor data fusion: A review of the state-of-the-art (vol 14, pg 28, 2013)," *Information Fusion,* vol. 14, no. 4, pp. 562-562, Oct 2013, doi: 10.1016/j.inffus.2012.10.004.
[64] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the Ieee,* vol. 85, no. 1, pp. 6-23, Jan 1997, doi: 10.1109/5.554205.
[65] F. E. White, "Data fusion lexicon," Joint Directors of Labs Washington DC, 1991.
[66] I. Pires, N. Garcia, N. Pombo, and F. Flórez-Revuelta, "From data acquisition to data fusion: a comprehensive review and a roadmap for the identification of activities of daily living using mobile devices," *Sensors,* vol. 16, no. 2, p. 184, 2016.
[67] M. Ribo, A. Pinz, and A. L. Fuhrmann, "A new optical tracking system for virtual and augmented reality applications," in *IMTC 2001. Proceedings of the 18th IEEE Instrumentation and Measurement Technology Conference. Rediscovering Measurement in the Age of Informatics (Cat. No. 01CH 37188)*, 2001, vol. 3: IEEE, pp. 1932-1936.
[68] C. Hillmann, "Comparing the Gear VR, Oculus Go, and Oculus Quest," in *Unreal for Mobile and Standalone VR*: Springer, 2019, pp. 141-167.
[69] R. Atac and E. Foxlin, "Scorpion Hybrid Optical-based Inertial Tracker (HObIT)," in *Conference on Head- and Helmet-Mounted Displays XVIII - Design and Applications*, Baltimore, MD, May 01 2013, vol. 8735, in Proceedings of SPIE, 2013, doi: 10.1117/12.2012194.
[70] A. Atrsaei, H. Salarieh, and A. Alasty, "Human Arm Motion Tracking by Orientation-Based Fusion of Inertial Sensors and Kinect Using Unscented Kalman Filter," *Journal of Biomechanical Engineering,* vol. 138, no. 9, pp. 091005-091005-13, 2016, doi: 10.1115/1.4034170.
[71] T. B. Rodrigues, C. Ó Catháin, D. M. Devine, K. Moran, N. E. O'Connor, and N. Murray, "An evaluation of a 3D multimodal marker-less motion analysis system," in *Proceedings of ACM Multimedia System Conference.*, 2019: ACM.
[72] W. Fang, L. Zheng, H. Deng, and H. Zhang, "Real-Time Motion Tracking for Mobile Augmented/Virtual Reality Using Adaptive Visual-Inertial Fusion," *Sensors,* vol. 17, no. 5, p. 1037, 2017.
[73] E. Foxlin and L. Naimark, "VIS-Tracker: A wearable vision-inertial self-tracker," in *IEEE Virtual Reality 2003 Conference*, Los Angeles, Ca, Mar 22-26 2003, in Proceedings

of the IEEE Virtual Reality Annual International Symposium, 2003, pp. 199-206, doi: 10.1109/vr.2003.1191139. [Online]. Available: <Go to ISI>://WOS:000182252300026
[74] T. Calloway, D. B. Megherbi, and H. Zhang, "Global localization and tracking for wearable augmented reality in urban environments," in *2017 IEEE International Conference on Computational Intelligence and Virtual Environments for Measurement Systems and Applications (CIVEMSA)*, 26-28 June 2017 2017, pp. 105-110, doi: 10.1109/CIVEMSA.2017.7995310.
[75] E. Foxlin, T. Calloway, and H. Zhang, "Improved registration for vehicular AR using auto-harmonization," in *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 10-12 Sept. 2014 2014, pp. 105-112, doi: 10.1109/ISMAR.2014.6948415.
[76] A. Inc. "Antilatency." https://antilatency.com/ (accessed 12.09.2019, 2019).
[77] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., 2006.
[78] P. Gill and T. Vogelsang, "Lensless Smart Sensors: Optical and thermal sensing for the Internet of Things," in *2016 IEEE Symposium on VLSI Circuits (VLSI-Circuits)*, 15-17 June 2016 2016, pp. 1-2, doi: 10.1109/VLSIC.2016.7573486.
[79] C. Sun, "Shrinking the camera size," *Nature Materials,* vol. 16, p. 11, 12/20/online 2016, doi: 10.1038/nmat4833.
[80] A. Betancourt, P. Morerio, C. S. Regazzoni, and M. Rauterberg, "The evolution of first person vision methods: A survey," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 25, no. 5, pp. 744-760, 2015.
[81] R. Thabet, R. Mahmoudi, and M. H. Bedoui, "Image processing on mobile devices: An overview," in *International Image Processing, Applications and Systems Conference*, 5-7 Nov. 2014 2014, pp. 1-8, doi: 10.1109/IPAS.2014.7043267.
[82] A. Medien, "Implementation of a Low Cost Marker Based Infrared Optical Tracking System," 2006.
[83] F. Heide, M. Rouf, M. B. Hullin, B. Labitzke, W. Heidrich, and A. Kolb, "High-quality computational imaging through simple lenses," *ACM Trans. Graph.,* vol. 32, no. 5, pp. 1-14, 2013, doi: 10.1145/2516971.2516974.
[84] A. Small and S. Stahlheber, "Fluorophore localization algorithms for super-resolution microscopy," *Nat Meth,* Review vol. 11, no. 3, pp. 267-279, 03//print 2014, doi: 10.1038/nmeth.2844.
[85] R. B. Fisher and D. K. Naidu, "A Comparison of Algorithms for Subpixel Peak Detection," in *Image Technology: Advances in Image Processing, Multimedia and Machine Vision*, J. L. C. Sanz Ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 1996, pp. 385-404.
[86] R. Parthasarathy, "Rapid, accurate particle tracking by calculation of radial symmetry centers," *Nature Methods,* vol. 9, p. 724, 06/10/online 2012, doi: 10.1038/nmeth.2071.
[87] D. G. Bailey, "Sub-pixel estimation of local extrema."
[88] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Süsstrunk, "SLIC Superpixels Compared to State-of-the-Art Superpixel Methods," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, no. 11, pp. 2274-2282, 2012, doi: 10.1109/TPAMI.2012.120.
[89] D. Yao and L. Yu, "Sub-pixel interpolation of MPEG-4 motion compensation and its hardware implementation," vol. 39, pp. 1703-1707, 11/01 2005.
[90] A. Kaw and E. E. Kalu, *Numerical Methods with Applications: Abridged*. Lulu.com, 2008.
[91] M. P. Wilk, A. Urru, S. Tedesco, and B. O. Flynn, "Sub-pixel point detection algorithm for point tracking with low-power wearable camera systems: A simplified linear

interpolation," in *2017 28th Irish Signals and Systems Conference (ISSC)*, 20-21 June 2017 2017, pp. 1-6, doi: 10.1109/ISSC.2017.7983629.
[92] I. OmniVision Technologies. "OV8865." http://www.ovt.com/sensors/OV8865 (accessed 23.02.2018, 2018).
[93] I. Microsoft. "Microsoft Surface Pro 4, 6th Gen i5." https://surfacetip.com/surface-pro-2017-vs-surface-pro-4/ (accessed 23.02.2018, 2018).
[94] E. O. Inc. "1" Diameter, Optical Cast Plastic IR Longpass Filter." Edmund Optics Inc. https://www.edmundoptics.com/optics/optical-filters/longpass-edge-filters/1quot-diameter-optical-cast-plastic-ir-longpass-filter/ (accessed 19.02.2018, 2018).
[95] S. D. Limited. "Bluetooth V4.0 HM-11 BLE Module." http://wiki.seeed.cc/Bluetooth_V4.0_HM_11_BLE_Module/ (accessed 23.02.2018, 2018).
[96] I. Vishay Intertechnology. "High Speed Infrared Emitting Diode, 940 nm, VSMB11940X01." http://www.farnell.com/datasheets/2245170.pdf?_ga=2.184115750.1623767235.1519051613-1677736993.1519051613&_gac=1.15708484.1519051613.Cj0KCQiAiKrUBRD6ARIsADS2OLlYcKGvXIt139L6uYIypwMFTqdQy6r-30rp9fSntK9OXyQTsLnW8b8aAm_VEALw_wcB (accessed 19.02.2018, 2018).
[97] A. Limited. "NUCLEO-F401RE." https://os.mbed.com/platforms/ST-Nucleo-F401RE/ (accessed 23.02.2018, 2018).
[98] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. 29, no. 6, pp. 1153-1160, 1981, doi: 10.1109/TASSP.1981.1163711.
[99] T. M. Lehmann, C. Gonner, and K. Spitzer, "Survey: interpolation methods in medical image processing," *IEEE Transactions on Medical Imaging,* vol. 18, no. 11, pp. 1049-1075, 1999, doi: 10.1109/42.816070.
[100] T. J. Atherton and D. J. Kerbyson, "Size invariant circle detection," *Image and Vision computing,* vol. 17, no. 11, pp. 795-803, 1999.
[101] M. P. Wilk and B. Q. Flynn, "Reference Point Estimation Technique for Direct Validation of Subpixel Point Detection Algorithms for Internet of Things," in *2019 30th Irish Signals and Systems Conference (ISSC)*, 17-18 June 2019 2019, pp. 1-5, doi: 10.1109/ISSC.2019.8904921.
[102] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern classification*. Wiley New York, 1973.
[103] W. De Vries, H. Veeger, C. Baten, and F. Van Der Helm, "Magnetic distortion in motion labs, implications for validating inertial magnetic sensors," *Gait & posture,* vol. 29, no. 4, pp. 535-541, 2009.
[104] R. Atac, S. Spink, T. Calloway, and E. Foxlin, "Scorpion Hybrid Optical based Inertial Tracker (HObIT) Test Results," in *Conference on Display Technologies and Applications for Defense, Security, and Avionics VIII*, Baltimore, MA, May 07-08 2014, vol. 9086, in Proceedings of SPIE, 2014, doi: 10.1117/12.2050363. [Online]. Available: <Go to ISI>://WOS:000343864900018
[105] 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.
[106] L. Abraham, A. Urru, M. P. Wilk, S. Tedesco, M. Walsh, and B. O'Flynn, "Point tracking with lensless smart sensors," in *SENSORS, 2017 IEEE*, 2017: IEEE, pp. 1-3.
[107] B. Siciliano and O. Khatib, *Springer handbook of robotics*. Springer, 2016.
[108] P. Sturm, "Pinhole Camera Model," in *Computer Vision: A Reference Guide*, K. Ikeuchi Ed. Boston, MA: Springer US, 2014, pp. 610-613.

[109] Z. Taylor. "Project 3D into 2D image coordinates using a camera model." MathWorks. https://uk.mathworks.com/matlabcentral/fileexchange/48752-project-3d-into-2d-image-coordinates-using-a-camera-model?s_tid=prof_contriblnk (accessed 31.08.2019, 2019).
[110] G. D. Myer *et al.*, "The back squat: A proposed assessment of functional deficits and technical factors that limit performance," *Strength and conditioning journal,* vol. 36, no. 6, p. 4, 2014.
[111] T. M. McLaughlin, T. J. Lardner, and C. J. Dillman, "Kinetics of the parallel squat," *Research Quarterly. American Alliance for Health, Physical Education and Recreation,* vol. 49, no. 2, pp. 175-189, 1978.
[112] K. Peterson, "Muscle Activation and Range of Motion Patterns of Individuals who Display a Lateral Hip Shift during an Overhead Squat," Masters Thesis, University of North Carolina at Chapel Hill Graduate School, 2019.
[113] P. Comfort and P. Kasim, "Optimizing squat technique," *Strength and Conditioning Journal,* vol. 29, no. 6, p. 10, 2007.
[114] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Transactions on Pattern Analysis & Machine Intelligence,* no. 10, pp. 965-980, 1992.
[115] C. Bishop and A. Turner, "Integrated Approach to Correcting the High-Bar Back Squat From "Excessive Forward Leaning"," *Strength & Conditioning Journal,* vol. 39, no. 6, pp. 46-53, 2017.
[116] R. F. Escamilla *et al.*, "Effects of technique variations on knee biomechanics during the squat and leg press," *Medicine & Science in Sports & Exercise,* vol. 33, no. 9, pp. 1552-1566, 2001.
[117] D. Diggin *et al.*, "A biomechanical analysis of front versus back squat: injury implications," in *ISBS-Conference Proceedings Archive*, 2011, vol. 1, no. 1.
[118] Manfrotto. "Manfrotto MN755XB Aluminium Video Tripod." https://www.barkerphotographic.ie/manfrotto-mn755xb-tripod-legs/camera-accessories/tripods (accessed 25.02.2019, 2019).
[119] Manfrotto. "Manfrotto 410 Junior Geared head." https://www.barkerphotographic.ie/manfrotto-410-junior-geared-head/camera-accessories (accessed 25.02.2019, 2019).
[120] "FIXKIT Mini Digital Inclinometer Angle Finder." Online. https://picclick.fr/FIXKIT-Mini-Digital-Inclinometer-Angle-Finder-Gauge-Level-142643236661.html (accessed 25.02.2019, 2019).
[121] "CPTCAM CP-80S Laser Distance Meter." https://es.tmart.com/CPTCAM-CP-80S-Portable-Handheld-80m-Mini-Laser-Rangefinder-Distance-Measuring-Meter-Black-Yellow_p338374.html (accessed 25.02.2019, 2019).
[122] Neewer. "Neewer 80 cm Aluminium Camera Track Slider Rail." https://www.amazon.co.uk/gp/product/B06Y46H989/ref=ppx_yo_dt_b_asin_title_o02_s02?ie=UTF8&psc=1 (accessed 9.08.2019, 2019).
[123] "NEMA 17 - 17HS4401 Stepper Motor." https://datasheet4u.com/datasheet-pdf-file/928661/MotionKing/17HS4401/1 (accessed 9.08.2019, 2019).
[124] "TB6600 Stepper Motor Driver." http://olimex.cl/website_MCI/static/documents/TB6600_data_sheet.pdf (accessed 9.08.2019, 2019).
[125] L. OpenMV. "OpenMV Cam H7." https://openmv.io/products/openmv-cam-h7 (accessed 13.09.2019, 2019).
[126] S. Microelectronics. "STM32H743VI." https://www.st.com/en/microcontrollers-microprocessors/stm32h743vi.html (accessed 13.09.2019, 2019).

[127] O. LLC. "Wifi Shield." https://openmv.io/collections/shields/products/wifi-shield (accessed 13.09.2019, 2019).
[128] MicroPython. "MicroPython." https://micropython.org/ (accessed 14.09.2019, 2019).
[129] U. Nations. "Ageing." https://www.un.org/en/sections/issues-depth/ageing/ (accessed 13.05.2019, 2019).
[130] NHS. "Physical activity guidelines for adults." https://www.nhs.uk/live-well/exercise/ (accessed 16.09.2019, 2019).