

https://dspace.org/

Feasibility of region of interest selection preprocessing using a multi-photodiode fingerprint-based visible light positioning system

Achari, Dipika

2024

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/45422

Achari, Dipika. (2024). Feasibility of region of interest selection preprocessing using a multi-photodiode fingerprint-based visible light positioning system [Masters dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/45422

Downloaded from a DSpace repository. https://dspace.org/

FEASIBILITY OF REGION OF INTEREST SELECTION PREPROCESSING USING A MULTI-PHOTODIODE FINGERPRINT-BASED VISIBLE LIGHT POSITIONING SYSTEM

Submitted by

Dipika Achari
Supervised by Professor Ling Cheng



SEPTEMBER 2024

A dissertation submitted to the Faculty of Engineering and the Built Environment, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Engineering.

ABSTRACT

This research presents a novel Multi-Photodiode Fingerprint-Based Visible Light Positioning (VLP) system aimed at improving the accuracy and reducing the computational expenses of indoor localization. The system leverages an advanced K-Nearest Neighbors (KNN) algorithm, enhanced by Signal Strength Clustering, alongside a region selection strategy based on frequency-modulated VLC encoded IDs. Through extensive simulations, the system demonstrated a notable reduction in Mean Absolute Error (MAE) to approximately 2.5 meters, with a Root Mean Square Error (RMSE) of around 3.0 meters. In addition, the system exhibited robustness across varying ambient light conditions and room sizes, maintaining an accuracy rate of 95%, even in challenging environments. Analysis revealed that error rates increased in larger rooms, with average errors ranging from 1.50 meters in smaller spaces to 3.51 meters in larger environments. This suggests that while the system is effective in smaller areas, its accuracy diminishes slightly as room size expands. However, integrating frequency domain analysis and region of interest (ROI) selection proved to be a practical approach, enhancing the overall performance of the VLP system by providing faster and more accurate indoor navigation. Future research includes exploring advanced modulation techniques integrating supplementary sensing technologies and fine-tuning the algorithm parameters to improve the system's accuracy and reliability, especially in more complex or dynamic environments.

Keywords: Visible Light Positioning, Indoor Positioning, Multi-Photodiode, Fingerprint-Based, K-Nearest Neighbors (KNN), Modulation Techniques, Signal Strength Clustering, Region of Interest Pre-Processing, System Optimization, Indoor Navigation

“No Victory Without Suffering”— J.R.R. Tolkien

DECLARATION

Declaration of Authorship I, DIPIKA ACHARI, hereby declare that the dissertation titled, "Feasibility of region of interest selection preprocessing using a multi-photodiode fingerprint-based visible light positioning system," is the result of my independent effort, with no external assistance except where duly acknowledged. This dissertation is being submitted in partial fulfillment of the requirements for the degree of Master of Science in Engineering at the University of the Witwatersrand, Johannesburg. It has not been previously submitted for any degree or examination at any other academic institution.

Signed: Johannesburg

Date: 08 September, 2024

Signature: 

Name: Dipika Achari

Student Number: 442344

ACKNOWLEDGMENTS

I am deeply grateful to Sri Pratyangira Devi and my spiritual guides, whose unwavering support has been instrumental in helping me complete this journey. Their wisdom and encouragement have given me the strength and courage to push beyond my limits and achieve what I once thought was impossible.

I want to express my heartfelt appreciation to Professor Ling Cheng, whose guidance and dedication have inspired me throughout this research. Despite the challenges and setbacks, you never lost faith in me and supported and motivated me to keep going. Your commitment to my success has been invaluable, and I cannot thank you enough for your efforts.

I also want to thank my dear husband, Shamin, whose love, support, and encouragement have been my anchor throughout this journey. Your unwavering belief in me, even when I doubted myself, has given me the strength and determination to keep moving forward. Your patience, compassion, and understanding have been invaluable, and I cannot imagine having gone through this without you by my side.

To my parents, Devindra and Yeshmin Harduth, I want to express my sincere appreciation for your unwavering support and love. Your constant check-ins, advice, and encouragement have been a source of comfort and inspiration, and I am immensely grateful for the sacrifices you have made to help me achieve my dreams.

I also want to thank my in-laws, Gonseelan and Roshni Acahri, for their continuous support and understanding. Your unwavering faith in our aspirations and willingness to support us in pursuing them has been remarkable, and I consider myself lucky to have such loving and supportive in-laws.

Finally, thanks to the University of Witwatersrand and the School of Electrical and Information Engineering for their support and funding, which made this research possible. To everyone who has contributed in various ways to this research, I want to thank you for your support and for joining me on this journey. Your contributions have significantly impacted the outcome of this work, and I will always cherish and remember them.

Thank you.

"Deeds will not be less valiant because they are upraised." J. R. R. Tolkien

TABLE OF CONTENTS

ACKNOWLEDGEMENTS **5**

LIST OF FIGURES **i**

LIST OF TABLES **ii**

LIST OF ABBREVIATIONS **iii**

1 INTRODUCTION **1**

1.1 Document Organization 1

1.2 Motivation 2

1.3 Research Questions 3

1.4 Research Objectives 3

1.5 Scope and limitations of the study 4

1.5.1 Scope 4

1.5.2 Limitations 5

1.6 Significance of the study 6

2 LITERATURE REVIEW **8**

2.1 Overview of Visible Light Positioning Systems 8

2.2 Key concepts and theories in VLP and indoor localisation 9

2.2.1 Fingerprinting Techniques in Indoor Localisation 9

2.2.2 Multiple Photodiode Fingerprinting for Visible Light Positioning 10

2.2.3 Pre-processing Techniques and Region of Interest Selection 11

2.3 Comparison with Existing Systems 12

2.4 Limitations and challenges in VLP systems 12

2.4.1 Line-of-Sight Dependency 12

2.4.2 Environmental Variability 12

2.4.3 Scalability and Deployment Complexity . . . 13
2.4.4 Interference, Cross-Talk, Security, and Privacy Concerns . . . 13
2.4.5 Mobility and Tracking . . . 14
2.5 Gap Analysis and Research Opportunities . . . 15

3 BACKGROUND 17

3.1 Background: Lambertian Model in Visible Light Positioning . . . 17
3.1.1 Characteristics of the Lambertian Model . . . 17
3.2 Signal Propagation in Indoor Environments for VLP . . . 18
3.2.1 Signal Degradation . . . 18
3.2.2 Multipath Reflections . . . 18
3.2.3 Noise and Interference . . . 19
3.2.4 Final Signal Model . . . 19
3.2.5 Application in Visible Light Positioning . . . 20
3.2.6 Advantages and Limitations . . . 20
3.3 Fingerprint Localisation Algorithm . . . 21
3.3.1 Basic Principle . . . 21
3.3.2 Signal Characteristics: . . . 22
3.3.3 Algorithm Steps . . . 22
3.3.4 Matching Techniques . . . 22
3.3.5 Challenges and Improvements . . . 24
3.3.6 K-Nearest Neighbors (KNN) Algorithm . . . 24
3.4 K-Nearest Neighbors (KNN) with Signal Strength Clustering . . . 29
3.4.1 Frequency Modulation in Visible Light Positioning Systems . . . 31

4 RESEARCH METHODOLOGY 34

4.1 Research Design . . . 34
4.2 Simulation Environment . . . 34

4.3 Algorithm Development 34
4.3.1 Online Phase 36
4.4 Performance Testing and Benchmarking 36
4.5 Data Analysis 37
4.6 LED Selection and Region Identification 38
4.6.1 Frequency Modulation for Signal Differentiation 39
4.6.2 Dynamic ID Assignment 39
4.6.3 Improved Area Identification 40
4.6.4 Pulse Amplitude Modulation (PAM) Modulation 40
4.6.5 Region of Interest Identification 41
4.6.6 Frequency Analysis 41
4.6.7 Preprocessing Algorithm 43
4.6.8 Integration with Fingerprinting Database 43
4.6.9 Simulation of Region Selection Preprocessing 43
4.6.10 Performance Metrics for Preprocessing 43
4.7 Determining Parameters in the WKNN Algorithm with Signal Clustering 44
4.7.1 Parameter Tuning for WKNN Algorithm with Signal Clustering 44
4.7.2 Estimate Positions with WKNN and Clustering 46
4.8 Methodology Summary 46

5 RESULTS AND ANALYSIS **47**

5.1 Computational Complexity Analysis 47
5.2 Results of VLP system with WKNN Without ROI Selection & Room size 4mx4m 49
5.3 Results of VLP system with WKNN with Signal Clustering Optimization and ROI Selection & Room Size 4mx4m 49
5.4 Overall System Performance 52
5.5 Comparison of Proposed Optimised VLP System Against Benchmark System 61
5.5.1 Analysis of Factors Contributing to Performance Difference Between The Two Systems 63

5.5.2 Computation Complexity Analysis Between The Two Systems 64

5.6 Summary of Findings 65

6 RECOMMENDATIONS AND CONCLUSION **67**

6.1 Recommendations 67

6.1.1 Conclusion 69

LIST OF FIGURES

3-1 Lambertian model 17

3-2 Fingerprint Localisation Algorithm Flowchart 23

3-3 K-Nearest Neighbors Algorithm 25

3-4 Diagram illustrating the K-Nearest Neighbors (KNN) with Signal Strength Clustering in VLP process 32

4-1 Proposed ROI & FingerPrintbased VLP Algorithm 38

4-2 Flow Diagram for LED ID Assignment Process 40

4-3 Example of LED Selection and Room Layout 42

5-1 System Complexity over Varying Room Sizes 48

5-2 Performance Metrics of VLP System Without ROI & Room Size 4mx4m 50

5-3 Performance Metrics of Optimized VLP System & Room Size 4mx4m 51

5-4 Room size: 4m x 4m, Avg. error: 1.50m 54

5-5 Room size: 7m x 7m, Avg. error: 1.66m 55

5-6 Room size: 13m x 13m, Avg. error: 1.99m 56

5-7 Room size: 17m x 17m, Avg. error: 2.21m 57

5-8 Room size: 45m x 45m, Avg. error: 3.51m 59

5-9 Room size: 32m x 32m, Avg. error: 3.03m 60

5-10 Scatter Plot of Errors with Trend Lines for Proposed VLP System Against Benchmark System 62

LIST OF TABLES

3.1 Distance Metrics 26

3.2 Comparative Analysis of Machine Learning Algorithms in VLP Systems 27

5.1 System Components and Their Complexities 47

5.2 Average Performance Metrics of The System Without ROI & Room Size 4mx4m 49

5.3 Average Performance Metrics of Optimized VLP System & Room Size 4mx4m 50

5.4 Errors of Optimized VLP System & Room Size 4mx4m under Higher Ambient Light Conditions 51

5.5 Average Summarised Performance Metrics of the VLP System 53

5.6 Summary of Performance Metrics for Proposed VLP System Against Benchmark System . . 62

LIST OF SYMBOLS AND ABBREVIATIONS

Term	Abbreviation
Visible Light Positioning	VLP
Visible Light Communication	VLC
Light Emitting Diode	LED
Data Base	db
PhotoDiode	PD
Weighted K-nearest neighbour	WKNN
K-nearest neighbour	KNN
Received Signal Strength	RSS
Three Dimensional	3-D
Deep Neural network	DNN
Extreme Learning Machine	ELM
Optical Camera Communication	OCC
Region of Interest	ROI
Big O notation-complexity of an algorithm in terms of input size	O
size of the input or the number of elements	N
number of measurements for fingerprint db or the number of LEDs	M
number of Queries	Q
Clusters or number of Clusters	C
light intensity at an angle	$I(\theta)$
angle from the normal	θ
peak intensity at normal emission	I_0
represents the power density at a	I_0
given distance from the light source	d
emitter's physical area	A
Lambertian emission order	m
LED's semi-angle at half power	ϕ
optical filter's gain	$T(\psi)$
optical concentrator's gain	$g(\psi)$
the angle of incidence with respect to the receiver	ψ

1. INTRODUCTION

Indoor localisation has been a long-standing challenge in navigation and object tracking, as traditional technologies like radar-based air traffic control, sonar, and GPS are not always effective in indoor environments due to infrastructure limitations, electromagnetic interference, and signal attenuation [4, 21]. While alternative technologies like wireless local area networks, Bluetooth, radio frequency identification, and ultra-wideband Bluetooth have been studied, they often require costly infrastructure upgrades and still struggle with signal interference issues [16, 21].

Fortunately, advancements in Visible Light Positioning (VLP) systems have emerged as a promising solution for indoor localisation challenges [27, 29]. VLP leverages visible light communication (VLC) technology, which uses light-emitting diodes (LEDs) and photodiodes (PDs) to operate on a line-of-sight principle, minimising interference and enhancing security [8, 39]. As a result, VLP technology has gained prominence in applications like robotics, drone tracking, warehouse management, and mining [58, 63].

This research introduces a VLP system that employs multiple photodiodes and incorporates advanced pre-processing techniques to predict precise positions [5, 46]. We conduct simulations to assess the system's feasibility and reduce practical implementation complexity [32, 57]. The primary contributions of this research are the proposal of a multiple photodiode-based VLP system with a fingerprint database, the utilisation of the Weighted K-nearest neighbour (WKNN) algorithm, and the introduction of a region selection algorithm employing frequency domain analysis to identify dominant frequencies and corresponding quadrants for fingerprint database reduction and complexity reduction [34, 52].

1.1 Document Organization

The remainder of this document is organised as follows:

The motivation for conducting research is explained in section 1.2. Section 1.3 outlines the research questions and objectives, and Chapter 2 provides an overview and comparison of existing VLP systems. Chapter 3 presents background information on applied techniques. While Chapter 4 details the methodology. Results analysis are presented in Chapter 5, and finally, Chapter 6 gives possible recommendations for

improving the system and concludes the study.

1.2 Motivation

The demand for accurate and dependable indoor positioning systems has rapidly increased across various industries, including intelligent buildings, healthcare facilities, logistics, and retail environments [5, 39]. However, traditional positioning technologies, such as GPS, have limitations in indoor environments due to signal attenuation and multi-path interference [27, 32]. Therefore, there is an urgent need for cost-effective, innovative solutions capable of providing precise indoor localisation capabilities.

One promising alternative that has emerged is visible light positioning systems [16, 21]. These systems use Light-Emitting Diodes (LEDs) to convey data and estimate the locations of receivers, which can provide opportunities for indoor navigation, asset tracking, and location-based services [8, 48]. Despite the advancements in visible light positioning, challenges like accuracy, robustness, and scalability persist [58, 63].

One of this field's most intriguing research areas is multiple photodiode fingerprint-based positioning [5, 46]. This approach can potentially enhance the performance of visible light positioning systems. However, there is a critical research gap in the literature regarding the feasibility and efficacy of this system in real-world scenarios [52, 57].

Therefore, this research aims to systematically evaluate the multiple photodiode fingerprints visible light positioning system's performance in various environmental conditions [29, 34]. This study aims to contribute to the progression of visual light positioning technology and provide valuable insights into its practical implementation.

The findings from this research have significant implications for industries and sectors that require precise indoor localisation. This technology can help optimise operations, improve safety and security measures, and provide customised services anchored on accurate location data. Furthermore, these findings can pave the way for future visible light positioning systems, overcoming existing limitations and unleashing their full potential.

In summary, this research is motivated by the need for enhanced indoor positioning systems and the potential of visible light positioning technology to fulfil this need. Through a thorough evaluation of the

multiple photodiode fingerprint visible light positioning system, this research aspires to advance the field of indoor localisation and enable its practical integration across various industries and domains.

1.3 Research Questions

The research questions addressed in this dissertation are as follows:

1. What methods of VLC modulation can be employed to enhance positioning accuracy in multiple photodiode VLP systems when compared to traditional square-wave modulation [27, 57]?
2. How do different distance metrics, such as Manhattan and chi-square chord distances, impact the complexity and accuracy of multi-PD VLP systems [5, 46]?
3. Can regression calibration effectively reduce the size of the fingerprint database, thereby minimising the number of offline readings required for accurate positioning in VLP systems [58, 63]?
4. To what extent does region selection pre-processing affect computation time and system complexity in multiple photodiode VLP systems [34, 52]?
5. Is frequency-based ID feature selection pre-processing feasible to maintain accuracy and reduce complexity as the VLP system scales up [16, 29]?

1.4 Research Objectives

The objectives of this research are as follows:

1. Investigate various VLC modulation methods to identify those significantly improving positioning accuracy in multiple photodiode VLP systems compared to square-wave modulation [27, 57].
2. Implement a simulation of a fingerprint-based VLP system using the Weighted K-nearest neighbour (WKNN) algorithm for precise position determination [5, 46].
3. Examine the effects of using different distance metrics (Manhattan and chi-square chord distances) on the complexity and accuracy of multiple photodiode VLP systems [58, 63].

4. Explore regression calibration techniques to decrease the fingerprint database size, aiming to minimise the number of offline readings needed for accurate positioning [34, 52].

5. Evaluate the impact of region selection pre-processing on computation time and system complexity in multiple photodiode VLP systems [16, 29].

6. Assess the feasibility of frequency-based ID feature selection pre-processing as the VLP system scales up, focusing on maintaining accuracy while reducing system complexity [8, 48].

1.5 Scope and limitations of the study

This research aims to design, develop, and rigorously evaluate a visible light positioning system utilising multiple photodiodes. The focus will be on fingerprint-based techniques, including using a fingerprint database with frequency-based ID pre-processing and region of interest selection for indoor localisation.

1.5.1 Scope

1. Comparative Analysis: The study will include a comprehensive comparative analysis of the proposed multiple photodiode system against visible light positioning technologies. Evaluation metrics will encompass computational time complexity, fingerprint database reduction strategies, algorithmic complexity, and positioning accuracy.

2. Simulation-Based Approach: Given the complexity of the proposed system, the study will rely on simulations to explore its performance. The simulations will be conducted across various scenarios, considering room sizes, LED configurations, and environmental conditions.

3. Investigation of Advanced Techniques: The research will explore advanced visible light communication modulation methods to enhance positioning accuracy, including methods beyond square-wave modulation. Additionally, the study will employ the Weighted K-nearest neighbour (WKNN) algorithm for position determination.

1.5.2 Limitations

1. Simulation-Based Evaluation: The research is limited to simulation-based evaluations and does not involve practical implementation. Consequently, real-world performance, practical challenges, and unforeseen issues may still need to be represented.

2. Parameter-Specific Findings: The proposed system will be evaluated within predefined parameters, including specific room sizes and LED configurations. The generalizability of results to diverse environments and setups may be constrained.

3. Focus on Photodiodes and Visible Light Communication: This study concentrates exclusively on using photodiodes and visible light communication for indoor localisation. Alternative localisation technologies and sensor combinations are not within the scope.

4. Line-of-Sight Assumption: The proposed system's effectiveness relies on the assumption of unobstructed line-of-sight communication, which may limit its practicality in environments with substantial obstacles or interference.

5. Emphasis on Technical Aspects: While the research prioritises accuracy and complexity aspects of the system, other vital factors such as power efficiency, cost-effectiveness, and scalability may receive limited attention

6. Exclusion of Security and Privacy Analysis: This study does not address potential security or privacy concerns arising from using visible light communication and fingerprint databases for indoor localisation. While these issues are important, especially in sensitive environments, they are beyond the scope of the current research and represent an area for future investigation.

7. External Factors Not Exhaustively Explored: The performance of the proposed system may be influenced by external factors like ambient lighting conditions and noise levels, which may need to be comprehensively analysed in this study.

1.6 Significance of the study

This research addresses critical challenges and unlocks innovative opportunities in indoor positioning systems. Various pivotal factors support the importance of this research.

1. Enhancing Indoor Localisation Precision: The research aims to significantly enhance the accuracy of indoor positioning, which is a fundamental requirement in various sectors, including intelligent buildings, healthcare facilities, logistics, and retail environments. This research seeks to deliver precise location information by leveraging advanced techniques and multiple photodiodes, revolutionising how indoor spaces are navigated and managed [23].

2. Overcoming Limitations of Existing Technologies: Traditional indoor positioning methods like GPS fall short indoors due to signal attenuation and multi-path interference. The proposed multiple photodiode fingerprint visible light positioning system offers a promising alternative. By mitigating these limitations, it has the potential to fill a significant technology gap and enable reliable indoor localisation [6].

3. Context-Aware Applications: Beyond accurate positioning, this research paves the way for context-aware applications. The system's ability to determine location and contextual information opens doors to personalised indoor navigation, asset tracking, and location-based services [47]. These applications have far-reaching implications across industries, enhancing efficiency, safety, and user experience.

4. Operational Optimisation: The findings of this research hold practical implications for a wide array of industries, enabling organisations to optimise their operations. Whether streamlining logistics, improving healthcare workflows, enhancing security measures, or delivering tailored services, the precise location information derived from this system can lead to significant operational improvements [18].

5. Future Technology Advancements: This study advances visible light positioning technology. By rigorously evaluating the feasibility and effectiveness of multiple photodiode fingerprinting, it addresses

current challenges and guides the design and development of future systems [40]. This research is, therefore, a building block toward unlocking the full potential of visible light positioning.

This research signifies a small effort to revolutionise indoor positioning, offering solutions to existing limitations, enabling context-aware applications, and driving operational efficiency across industries. Its contributions extend beyond the current state of technology, shaping the future of indoor localisation and its practical implementation.

2. LITERATURE REVIEW

This literature review explores the recent advancements in visible light positioning systems, focusing on fingerprinting for indoor localisation. The need for accurate and versatile indoor positioning solutions has grown significantly over the years, and visible light positioning has emerged as a promising alternative to overcome the limitations of conventional positioning systems by leveraging the existing lighting infrastructure. The review aims to provide a comprehensive overview of the research conducted in the field, highlighting the various methodologies, approaches, and persistent challenges associated with visible light positioning. By scrutinising relevant studies, it aims to establish a foundation for a systematic assessment of the feasibility and effectiveness of multiple photodiode fingerprint-based visible light positioning systems. The subsequent sections will navigate the existing research, exploring the intricacies of modulation techniques, distance metrics, computational complexities, and other facets constituting the framework of visible light positioning. The primary goal is to gain a profound understanding of visible light positioning and identify unexplored research areas.

2.1 Overview of Visible Light Positioning Systems

VLP is a promising technology for indoor positioning. Unlike traditional localisation systems, it uses light to locate a device's position. VLP systems infer the position of a device using the characteristics of the light source within an area or by encoding the light source and using VLC (visible light communication) with data or an ID. A receiver, typically a camera or photodiode, then captures these characteristics or data online and offline. The offline phase calibrates the system and creates a map of the light characteristics and positions. When looking at a VLP system, the position (x,y) has a corresponding vector of either VLC data or RSS values. The unique ID or characteristics mapped are then compared to the real-time readings and are used to determine the position of the device or receiver [23]. This section overviews visible light positioning systems' fundamental principles and components.

2.2 Key concepts and theories in VLP and indoor localisation

Visible Light Positioning (VLP) and indoor localisation are based on fundamental concepts and theories used to realise accurate positioning and tracking. The following sections outline these primary concepts and theories.

2.2.1 Fingerprinting Techniques in Indoor Localisation

The term "fingerprinting" originated from comparing specific features of an object (originally a human fingerprint) with those in a database to identify a match. Although the term is now used in various contexts, the basic principle remains the same.

Indoor Visible Light Positioning (VLP) systems often employ fingerprinting techniques to gain precise localisation. This approach involves collecting signal measurements, such as Received Signal Strength (RSS) and signal propagation time, at various locations within the indoor environment. These measurements are then catalogued to establish a comprehensive database of reference 'fingerprints', each corresponding to a specific indoor location.

The fingerprint database is a critical resource for the VLP system, serving as a reference library against which real-time signal measurements are compared to accurately determine the receiver's position. Fingerprinting techniques offer several advantages, including, in some cases, high precision, robustness to environmental changes in practical systems, and context awareness. However, they also face resource-intensive database construction in both simulations and practical implementation and signal variability challenges in practical implementation. Research has shown the effectiveness of fingerprinting techniques in indoor VLP localisation, highlighting their feasibility in achieving precise positioning with minimal training time. For example, Y. Chen, W. Guan, J. Li, and H. Song [48] introduced a real-time 3-D VLP system that employed fingerprinting in combination with Extreme Learning Machine (ELM) algorithms. Their study demonstrated that fingerprinting techniques could achieve accurate positioning with reduced training time. This work highlights the potential of combining advanced machine learning techniques with traditional fingerprinting to overcome some of the inherent challenges [48].

2.2.2 Multiple Photodiode Fingerprinting for Visible Light Positioning

Using multiple photodiodes improves the accuracy and reliability of VLP systems. By incorporating many photodiodes, spatial resolution is significantly enhanced. For instance, T. W. Fakhrul Alam, Moi Tin Chew, and G. S. Gupta's system demonstrated the effectiveness of triangulation in achieving precise receiver positioning within an illuminated area [6]. This system accurately determines the receiver's position by calculating the intersection of signals received from different angles.

A key benefit of using multiple photodiodes is their potential ability to mitigate multi-path interference, a common issue in realistic indoor environments where light signals reflect off surfaces. By analysing the differences in the signals received by various photodiodes, these systems can filter out erroneous reflections and focus on the direct signal path, leading to more accurate positioning [2, 55].

Additionally, the redundancy provided by multiple photodiodes contributes to more excellent system reliability. In scenarios where signal degradation or loss might occur—due to obstacles or poor signal conditions—the presence of multiple photodiodes ensures that the system can still maintain signal integrity by receiving signals from alternate paths. This makes the system less vulnerable to disruptions, especially in complex environments with occlusions [2, 55].

In such challenging environments, as addressed by Y. Chen, W. Guan, J. Li, and H. Song [48], systems with multiple photodiodes maintain accuracy by detecting signals from various vantage points. This allows them to adapt to obstacles and environmental changes while providing precise localisation.

Furthermore, the robustness of systems using multiple photodiodes extends to their ability to handle signal variability caused by fluctuating lighting conditions. By intelligently combining information from numerous photodiodes, these systems can adjust to changes in ambient light or other interference, ensuring consistent performance over time.

Another advantage of using multiple photodiodes is scalability. For example, scalability enables the system to maintain high precision even as the coverage area expands in systems like that proposed by T. T. S. H. L.-M. A. B. Manh The Van and Nguyen Van Tuan [54]. As more photodiodes are integrated into the system, the accuracy and resolution continue to improve, allowing the VLP system to cover larger spaces without sacrificing performance.

2.2.3 Pre-processing Techniques and Region of Interest Selection

Pre-processing techniques and Region of Interest (ROI) selection are key factors in improving the performance and efficiency of Visible Light Positioning (VLP) systems. These techniques are crucial in optimising signal data before applying fingerprint-based localisation methods, leading to more efficient and accurate positioning.

In VLP systems, pre-processing techniques refine raw signal data and extract meaningful information that enhances positioning. For example, T. W. Fakhrul Alam, Moi Tin Chew, and G. S. Gupta's work [6] demonstrates frequency-based identification (ID) feature selection as a pre-processing method. This technique reduces the complexity of the system while maintaining high accuracy. The system can streamline the fingerprint database by selecting the most relevant frequency components of the received signals, ensuring that only the most critical data is processed. This reduces computational load and enhances the system's speed and accuracy.

Another important technique in VLP systems is Region of Interest (ROI) selection. ROI selection involves focusing on a specific area within the indoor environment, allowing the system to concentrate its computational resources on that particular region. This approach is particularly valuable in large and complex indoor spaces, where processing every possible location would be inefficient. For instance, the study by A. Hasan, T. Glass, F. Alam, and M. Legg [18] illustrates the effective use of ROI selection in real-time 3-D VLP systems. Limiting the processing effort to the relevant spatial region significantly reduces the overall computational burden while maintaining precise positioning.

Combining pre-processing techniques, such as frequency-based ID selection, with ROI selection enhances the overall performance of VLP systems. These techniques work in tandem to optimise the system's accuracy and efficiency, ensuring that real-time positioning is fast and reliable. Furthermore, ROI selection minimises unnecessary complexity by focusing on the most relevant areas, which can significantly improve processing time and energy consumption.

In summary, integrating advanced pre-processing methods and targeted region selection is critical for enhancing the functionality of VLP systems. These approaches allow the system to operate efficiently without compromising precision, making them essential for modern VLP applications.

2.3 Comparison with Existing Systems

Benchmarking the proposed multiple photodiode fingerprint visible light positioning system against existing systems is crucial to assess its performance and potential advantages. This section reviews the literature to identify comparable systems and evaluates their computational time complexity, fingerprint database reduction techniques, and algorithm complexity vs. accuracy trade-offs. The comparison aims to highlight the novelty and potential contributions of the proposed system in the context of existing solutions.

2.4 Limitations and challenges in VLP systems

Visible Light Positioning (VLP) systems have gained considerable attention as a promising solution for indoor localisation. However, developing robust and adaptable VLP systems is complex and presents numerous challenges. Various studies have addressed several limitations, but many areas still require further exploration.

2.4.1 Line-of-Sight Dependency

A key limitation of many VLP systems is their dependence on line-of-sight (LOS) communication, which restricts the system's effectiveness in environments with obstacles or interference. Since most VLP systems rely exclusively on visible light communication without incorporating other localisation technologies or sensors, they may struggle in complex or dynamic indoor settings. This reliance on LOS can lead to significant performance degradation in obstructed environments, highlighting the need for more versatile systems to integrate additional technologies to enhance robustness and accuracy.

2.4.2 Environmental Variability

Another major challenge is generalising VLP systems across different indoor environments. Most studies evaluate or simulate these systems under specific conditions—such as controlled room sizes and LED configurations—that may not reflect the variability found in real-world settings. Factors such as ambient lighting, noise, and interference can significantly affect the performance of VLP systems but are often insufficiently

addressed in existing research. The adaptability of VLP technology to diverse environments and lighting conditions remains a crucial area for further investigation.

2.4.3 Scalability and Deployment Complexity

Many studies primarily focus on the technical aspects of VLP systems, such as accuracy and computational complexity, but fail to address broader factors critical for practical deployment. These factors include power consumption, cost-effectiveness, and scalability, essential for real-world applications. Deploying VLP systems in large-scale environments or across multiple buildings introduces additional challenges, such as the need for extensive infrastructure, maintenance, and optimisation. Moreover, the rapid pace of technological advancement makes it difficult for current systems to stay up-to-date with the latest innovations, requiring continuous improvements to ensure scalability and cost-efficiency.

Additionally, reliance on simulation-based evaluations is a significant limitation. While simulations offer controlled environments for testing, they may not fully capture the complexities of real-world scenarios, such as interference, varying lighting conditions, or deployment challenges. Practical implementations and field tests are necessary to address these gaps and validate the performance of VLP systems under real-world conditions.

2.4.4 Interference, Cross-Talk, Security, and Privacy Concerns

Additional limitations are related to the lack of discussion on security and privacy concerns associated with visible light communication and the use of fingerprint databases for indoor localisation. In sensitive environments such as healthcare or government facilities, security issues related to data transmission, interference between light sources, and unauthorised access to location data are crucial. Addressing these concerns through encryption, secure data storage, and privacy-preserving protocols is essential to facilitate the broader adoption of VLP systems.

2.4.5 Mobility and Tracking

Recent research has explored the dynamic localisation and tracking of mobile receivers in VLP systems, recognising the importance of mobility in real-world applications. For example, Chen et al. [48] introduced a real-time 3D VLP system with tracking capabilities, employing Extreme Learning Machine (ELM) technology to track mobile users within indoor environments dynamically. This capability is handy for applications such as asset tracking, indoor navigation, and context-aware services, where precise real-time monitoring is essential.

Clark et al. [14] investigated inter-mobile-device distance estimation for digital contact logging applications, primarily using Bluetooth technology. While not directly focused on VLP, their work demonstrates the potential of integrating mobile devices with VLP systems for tasks like contact tracing, social distancing enforcement, and location-based services. Such integration could significantly enhance the functionality and applicability of VLP systems in various contexts.

Aalimahmoodi et al. [1] explored how camera parameters impact Optical Camera Communication (OCC)-based indoor positioning systems. Although OCC is not strictly VLP, this research underscores the importance of optimising camera settings for tracking and localisation accuracy. Understanding these parameters can help improve VLP systems that rely on cameras for position estimation.

Loizos Kanaris et al. [22] proposed a hybrid system that combines radio fingerprinting with VLP for improved indoor localisation. Their sensor fusion approach—integrating data from multiple sources like VLP and radio signals—demonstrates how leveraging multiple sensors can enhance accuracy, robustness, and reliability in complex environments. This method highlights the potential of hybrid systems for dynamic tracking in scenarios where a single technology may not suffice.

Recent developments in machine learning have further enhanced dynamic position estimation in VLP systems. Liang et al.[29] proposed an inertial-aided VLP system that uses modulated LEDs and unmodulated lights as landmarks. The system can adapt to changing user positions over time by incorporating machine learning techniques, enabling more flexible and accurate tracking in dynamic environments.

2.5 Gap Analysis and Research Opportunities

After analysing the current research in visible light positioning (VLP) systems, several opportunities have emerged to advance knowledge in line with the objectives of this study. Previous studies, such as Alam et al. [6], underscore the importance of exploring advanced modulation techniques to improve positioning accuracy in multi-photodiode VLP systems. A critical gap exists in understanding the comparative effects of different distance metrics on the complexity and precision of these systems. Simon et al. [47] suggest that metrics like Manhattan and chi-square chord distances be studied to evaluate their influence on system performance. Additionally, regression calibration techniques specific to VLP systems could be optimised to reduce the number of offline readings necessary for accurate positioning, as highlighted by Chen et al. [27].

Enhancing region selection algorithms and addressing scalability issues are crucial for improving system efficiency. Zhang et al. [65] emphasise the need to assess how region selection pre-processing impacts computational time and system complexity. Furthermore, incorporating machine learning techniques—such as deep and reinforcement learning—could significantly enhance tracking accuracy and adaptability. Liang et al. [29] demonstrate that machine learning can be leveraged for dynamic position estimation. This opens the door to further research into real-time learning and system adaptability to evolving indoor conditions.

Another promising research direction involves multi-sensor fusion, as discussed by Loizos Kanaris et al. [22]. Future research could develop advanced fusion algorithms that integrate data from various sources, such as VLP, radio signals, and visual sensors, resulting in more accurate and robust tracking solutions. This is especially relevant in complex indoor environments that benefit from diverse sensor modalities.

Integrating mobile devices into VLP systems presents a promising opportunity, particularly for applications like contact tracing, social distancing, and location-based services. Future studies could explore methods to improve compatibility and tracking accuracy in mobile device integration. This may involve developing dedicated applications, protocols, or standards that facilitate seamless interaction between mobile devices and VLP infrastructure.

Finally, while several studies have explored dynamic localisation and tracking in VLP systems, further research is needed to refine and expand these capabilities. Future investigations could focus on developing

more robust tracking algorithms that account for challenging conditions such as rapid user movements, occlusions, or multi-path interference. Additionally, integrating inertial sensors (e.g., accelerometers and gyroscopes) further improves tracking accuracy, especially in scenarios where users' positions change rapidly.

In summary, the review of current VLP research reveals several pathways to improve the accuracy, efficiency, and robustness of VLP systems, paving the way for more advanced and versatile indoor localisation solutions.

3. BACKGROUND

3.1 Background: Lambertian Model in Visible Light Positioning

The Lambertian model is fundamental to Visible Light Positioning (VLP) systems. It provides insights into how light intensity distributes from LED sources to receivers in indoor environments. Characterising the optical power distribution, the model accurately estimates a receiver's position relative to the LED source, making it foundational in VLP system design. This section explores the Lambertian model's application, characteristics, benefits, and limitations within VLP systems, offering a comprehensive view of its importance in positioning and signal reception [20].

3.1.1 Characteristics of the Lambertian Model

The Lambertian model describes given by equation 3.1 the light intensity distribution from an LED based on the angle of emission. A fundamental assumption is that the light intensity is strongest along the normal to the surface of the LED and decreases as the angle between the direction of the light and the receiver increases, As shown in 3-1. This distribution is characterised by the Lambertian emission order,m, which defines how quickly the intensity falls off with the angle [23].

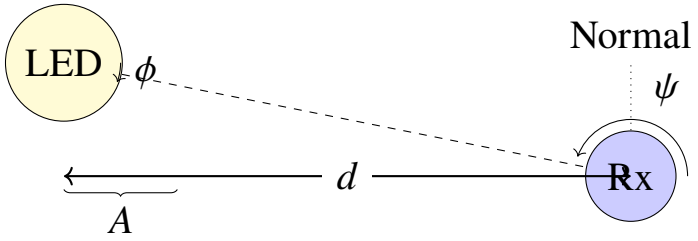


Figure 3-1: Lambertian model

$$I_0 = \frac{(m+1)A}{2\pi d^2} \cdot \frac{\cos^m(\phi)T(\psi)g(\psi)}{\cos(\psi)} \tag{3.1}$$

Here, I_0 represents the power density at a given distance d from the light source, A is the emitter's physical area, m is the Lambertian emission order related to the LED's semi-angle at half power ϕ, $T(\psi)$ is

the optical filter's gain, $g(\psi)$ is the optical concentrator's gain, and ψ is the angle of incidence with respect to the receiver's normal[23].

3.2 Signal Propagation in Indoor Environments for VLP

The propagation of light in an indoor environment for a Visible Light Positioning (VLP) system is primarily modelled by the Lambertian emission model, which describes the light intensity from an LED source at various angles. However, signal propagation in real environments is affected by several factors, including signal degradation and multipath reflections. These factors introduce complexities that need to be mathematically modelled.

3.2.1 Signal Degradation

Signal degradation due to distance is captured by the inverse square law. As the distance d increases, the received power P_{LED} decays according to:

$$P_{LED} \propto \frac{1}{d^2}$$

Additionally, signal degradation can be modelled as attenuation caused by environmental factors such as air particles, dust, and objects. This attenuation factor α can be added to the model:

$$P_{LED} = \frac{I_0 \cdot (m+1) \cdot A \cdot \cos^m(\theta) \cdot T(\psi) \cdot g(\psi) \cdot \cos(\psi)}{2\pi d^2} \cdot e^{-\alpha d}$$

Where α is the attenuation coefficient due to air or environmental obstacles.

3.2.2 Multipath Reflections

In real indoor environments, light may travel directly from the LED to the receiver and reflect off surfaces such as walls, floors, and ceilings. These reflections introduce multipath components, which can constructively or destructively interfere with the direct signal. The total received signal P_{total} is the sum of the direct path signal P_{direct} and the reflected path signals $P_{reflect}$.

For N reflective surfaces, the total received power can be expressed as:

$$P_{total} = P_{direct} + \sum_{i=1}^{N} P_{reflect,i}$$

Each reflected signal $P_{reflect,i}$ is modeled as:

$$P_{reflect,i} = \frac{\rho_i \cdot I_0 \cdot (m+1) \cdot A \cdot \cos^m(\theta_i) \cdot T(\psi_i) \cdot g(\psi_i) \cdot \cos(\psi_i)}{2\pi d_i^2} \cdot e^{-\alpha d_i}$$

Where:

- ρ_i is the reflectivity of surface i,
- θ_i is the angle of incidence of the light on surface i,
- d_i is the total travel distance of the light from the LED to the receiver after reflection from surface i,
- ψ_i is the angle of incidence of the reflected signal at the receiver.

3.2.3 Noise and Interference

The received signal is also affected by noise $N_{ambient}$, which can be modelled as Gaussian white noise. The total received signal strength is therefore:

$$P_{total} = P_{LED} + P_{reflect,i} + N_{ambient}$$

where $N_{ambient} \sim \mathcal{N}(\mu, \sigma^2)$, with μ representing the mean ambient light interference and σ^2 representing its variance.

3.2.4 Final Signal Model

Incorporating all these factors, the final signal received by the photodiode at the receiver can be written as:

$$P_{total} = \frac{I_0 \cdot (m+1) \cdot A \cdot \cos^m(\theta) \cdot T(\psi) \cdot g(\psi) \cdot \cos(\psi)}{2\pi d^2} \cdot e^{-\alpha d} + \sum_{i=1}^{N} P_{reflect,i} + N_{ambient}$$

This equation models the total signal at the receiver as a combination of direct light-reflected signals and ambient noise.

3.2.5 Application in Visible Light Positioning

The Lambertian model provides critical insights into how light intensity behaves. As the distance d from the LED increases, the intensity decreases following an inverse square law, which is crucial for estimating distances using trilateration or triangulation methods [23]. Additionally, the angle of emission psi plays a significant role, with intensity diminishing as the angle deviates from the normal (perpendicular) direction. This helps predict signal weakening at off-axis angles for accurate positioning. The Lambertian emission order m defines the concentration of light, where a higher m focuses light along the central axis and a lower m results in a more diffuse light spread, aiding in optimising system coverage and signal strength[23, 32]. This information is vital for accurately determining a receiver's position based on received signal strength. The model's key applications include:

- Designing systems for accurate indoor localisation in various applications such as retail environments and smart buildings.
- Providing a relatively simple and realistic representation of LED light distribution, enabling the implementation of straightforward algorithms for position estimation.

3.2.6 Advantages and Limitations

The Lambertian model offers several advantages and limitations in VLP systems: **Advantages:**

- Simplifies the representation of LED light distribution, facilitating the implementation of positioning algorithms. The model provides a straightforward and accurate way to estimate light distribution in indoor environments. This allows for efficient computation in real-time systems, making it ideal for VLP applications
- The model can be adjusted for different types of LEDs, allowing it to be used across various lighting systems.

Limitations:

- The Lambertian model assumes that the surface of the LED is a perfect diffuser, which is often not the case in real-world lighting systems. Many LED emitters have a non-ideal emission pattern that may deviate from the Lambertian assumption.
- The model does not account for environmental factors such as reflections, obstructions, or multipath interference, common in real-world indoor environments. These factors can cause deviations from the ideal Lambertian distribution.
- Additionally, accuracy decreases significantly with deviations in the receiver's angle of incidence or environments with reflective surfaces.

Despite its limitations, the Lambertian model remains popular in VLP systems due to its simplicity and practical trade-off between complexity and performance. The model provides sufficiently accurate localisation in controlled indoor environments for many practical applications. The Lambertian model's significance in VLP systems is well-documented in the literature [20, 25, 28, 37, 38, 41].

3.3 Fingerprint Localisation Algorithm

Fingerprint localisation algorithms are used in indoor positioning systems to estimate a device's location based on observed signal characteristics. These algorithms are often employed in environments where GPS signals are weak or unavailable.

3.3.1 Basic Principle

Fingerprint-based localisation starts by constructing a detailed database or map containing pre-recorded signal characteristics (e.g., Wi-Fi signal strength, Bluetooth signals, or other electromagnetic signals) at known reference points. During the positioning process, the system compares real-time observed signal characteristics with those in the database to estimate its position. This matching process can be done using pattern recognition algorithms, distance metrics, or machine learning models, with the accuracy depending on the granularity and quality of the fingerprint map [49].

3.3.2 Signal Characteristics:

The key signal characteristics for fingerprint localisation include the Received Signal Strength Indicator (RSSI), Channel State Information (CSI), and magnetic field strength. RSSI measures the power level of the received signals, providing a rough estimate of distance based on signal attenuation. Although RSSI is easy to collect, it is sensitive to interference and multipath effects, leading to inaccurate localisation in less controlled environments. CSI offers a more detailed insight into the signal's propagation characteristics, including phase and amplitude changes across multiple frequency channels, making it more robust against interference. However, it requires more complex hardware and computational resources. Magnetic field strength can also be used, as indoor environments have spatially varying magnetic fields due to the structural materials of buildings. Each of these signals varies uniquely at different locations, enabling the creation of a fingerprint map that can distinguish even closely spaced points [56].

3.3.3 Algorithm Steps

The algorithm for basic fingerprint localisation consists of two main phases. The process is illustrated in Diagram 3-2.

Offline Phase: Collect and store the signal characteristics at various known locations to create a fingerprint map.

Online Phase: Compare the real-time signal characteristics with the fingerprint map to estimate the device's location [19].

3.3.4 Matching Techniques

Several techniques exist for matching real-time (online) signal readings to the pre-collected database (offline readings). The most commonly used method is k-Nearest Neighbors (kNN). It is a simple and effective algorithm that identifies the k closest signal fingerprints from the database based on a predefined distance metric (e.g., Euclidean distance). The position is then estimated by averaging the coordinates of these k closest fingerprints. While kNN is computationally efficient and straightforward, its accuracy is highly dependent on

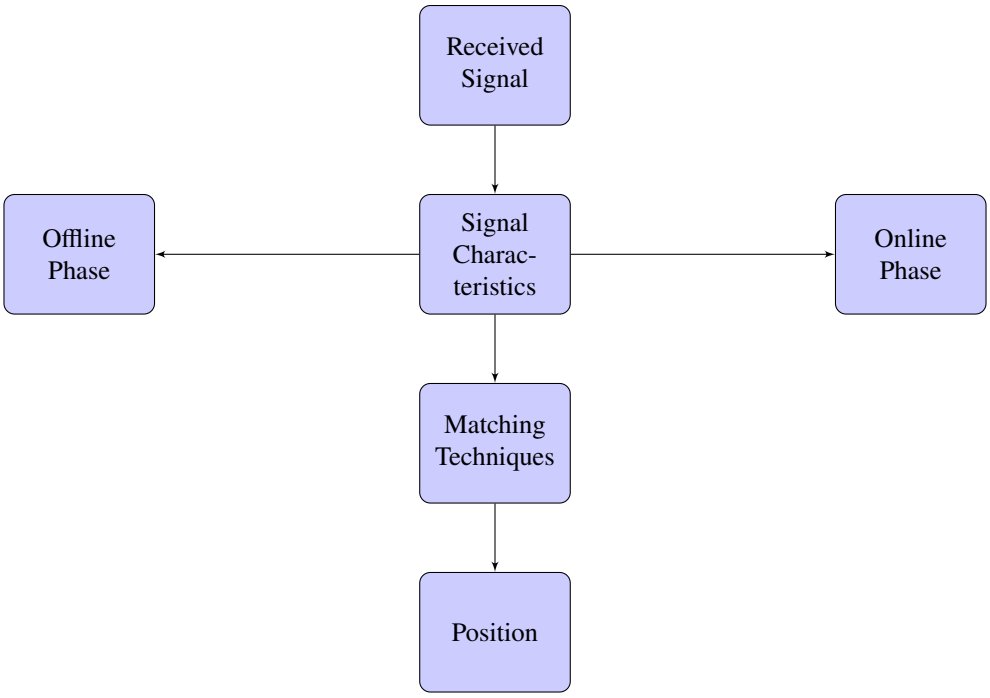


Figure 3-2: Fingerprint Localisation Algorithm Flowchart

the density of the fingerprint map and the choice of k. If k is too small, the system may be sensitive to noise, while a larger k can lead to errors due to averaging over distant points [19].

Advanced machine learning techniques, such as neural networks and support vector machines (SVMs), have been used for more accurate positioning. In particular, deep neural networks (DNNs) excel in non-linear pattern recognition, making them well-suited for indoor environments with complex signal propagation behaviours. These models can learn to extract relevant features from signal characteristics autonomously, enabling more precise location estimates even in environments with multipath interference or signal fluctuations [24].

Techniques like Multidimensional Scaling (MDS) are used to reduce the dimensionality of signal data without losing essential information. MDS transforms the high-dimensional signal space into a lower-dimensional representation, preserving relative distances between points. This dimensionality reduction technique speeds up the matching process and improves the system's scalability. MDS is used with large datasets or fingerprint maps, ensuring that the relative geometry of signal data is maintained [24].

3.3.5 Challenges and Improvements

The primary challenge in real-world fingerprint localisation is the dynamic nature of indoor environments. Changes such as furniture rearrangement, the introduction of obstacles, or even human movement can significantly alter signal propagation patterns. These variations lead to discrepancies between the fingerprint map (created during the initial offline phase) and the real-time signal readings, reducing the accuracy of the localisation system over time. As a result, frequent updates to the fingerprint map are necessary to ensure that the system reflects the current state of the environment. However, performing these updates manually is labour-intensive and time-consuming, particularly in large or highly dynamic spaces [15]. A possible solution is the development of automatic or semi-automatic fingerprint map update mechanisms, where machine learning algorithms are used to adjust the fingerprint database in response to environmental changes continuously. These algorithms can detect significant deviations in signal patterns and trigger updates, thus reducing the need for manual interventions. Additionally, methods such as crowd-sourced data collection leverage user devices to gather real-time signal readings, which are then aggregated and used to refine the fingerprint map dynamically. This approach minimises the cost of regular updates and enhances map accuracy by incorporating data from a broader range of environmental states[61].

3.3.6 K-Nearest Neighbors (KNN) Algorithm

KNN is a straightforward yet effective algorithm used in classification and regression tasks. It belongs to the instance-based learning category, often called lazy learning. Because it does not build an explicit model during the training phase. The KNN Algorithm stores all training data and performs computations only when making predictions. This "lazy" nature contrasts with 'eager' learning algorithms that construct a predictive model before receiving new data [26, 53].

3.3.6.1 Basic Principle

The Principle of KNN is that similar data points tend to be found near each other in feature space or that objects with similar characteristics are spatially close. This concept is rooted in the idea that data points belonging to the same class or having similar continuous values in regression tend to cluster together[26, 53].

3.3.6.2 Algorithm Steps

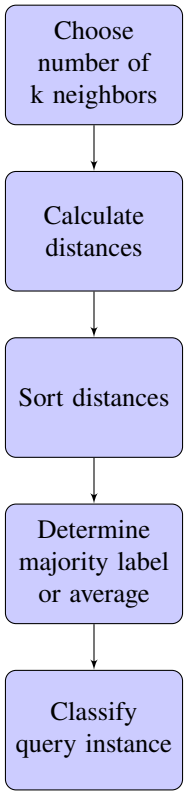


Figure 3-3: K-Nearest Neighbors Algorithm

The KNN algorithm involves several steps to classify or regress a query instance. Firstly, the value of K, representing the number of nearest neighbours to consider in the classification or regression process, is chosen. Then, the distance between the query instance and all the training samples in the dataset is calculated. Once the distances between the query instance and the training samples have been calculated, they are sorted in ascending order. The K closest points (i.e., the nearest neighbours) are selected based on this sorted list. KNN uses a majority voting mechanism for classification tasks, where the most frequent class label among the K nearest neighbours is assigned to the query point. In the regression case, KNN calculates the average value of the K neighbours to produce a continuous output. In both cases, each neighbour contributes equally to the final decision, although weighted KNN is a common variant where closer neighbours have more influence than distant ones. Finally, based on the majority label or average value, the algorithm classifies the query instance (for classification problems) or predicts a numerical value (for regression problems).

3.3.6.3 Distance Metrics

The algorithm computes the distance between the query instance and all the points in the training dataset. The most commonly used distance metric is Euclidean distance. Still, other metrics such as Manhattan distance, Minkowski distance, or Cosine similarity can be applied depending on the nature of the data and the problem at hand. For example, Manhattan distance may be preferred in grid-like environments, while Cosine similarity is often used in high-dimensional spaces. The choice of distance metric significantly influences the algorithm's performance. The table 3.1 gives the equations for a few distance metrics mentioned.

Table 3.1: Distance Metrics

Distance Metric	Equation
Euclidean Distance	$d(\mathbf{p}, \mathbf{q}) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}$
Manhattan Distance	$d(\mathbf{p}, \mathbf{q}) = \sum_{i=1}^{n} \lvert q_i - p_i \rvert$
Minkowski Distance	$d(\mathbf{p}, \mathbf{q}) = \left(\sum_{i=1}^{n} \lvert q_i - p_i \rvert^p\right)^{\frac{1}{p}}$

3.3.6.4 Choosing K

The choice of K can significantly impact the algorithm's performance. A small K value might lead to a model that is too sensitive to noise (overfitting), while a considerable K value could result in an overly general model (underfitting). Typically, K is chosen through cross-validation. Selecting the appropriate value for K is crucial. A small K value influences noise more in the result, while a considerable K value becomes computationally expensive. Cross-validation is commonly used to determine the optimal K value to find the optimal balance between bias and variance[31].

3.3.6.5 KNN compared to other Machine Learning Techniques

Machine learning techniques have been integrated with the KNN algorithm to enhance performance in various Visible Light Positioning (VLP) applications. Table 3.2 summarises various machine learning approaches applied to Visible Light Positioning (VLP) systems, showcasing high accuracy rates ranging from 90% to 96% and relatively low positioning errors (RMSE: 0.33m-0.52m, MAE: 0.23m-0.35m). Diverse algorithms are explored, including KNN, DNN, ANN, and hybrid approaches, with Deep Neural Networks and multiple-photodiode receiver techniques demonstrating promising results. The research focuses on improving accuracy, robustness, and adaptability in indoor environments, with innovative techniques such as adaptive

weighted KNN, tilted-angle insensitive RSS using DNNs, and modified path-loss models being developed. Overall, machine learning approaches consistently achieve high performance in VLP systems, with a trend toward hybrid techniques and adaptability to challenging conditions.

Table 3.2: Comparative Analysis of Machine Learning Algorithms in VLP Systems

No.	System Description	Algorithm	Acc (%)	RMSE (m)	MAE (m)	Key Findings	Ref.
1	Adaptive Weighted KNN Trilateration for VLP	KNN	92	0.48	0.32	Introduced adaptive weighted KNN for trilateration, enhancing accuracy	[55]
2	VLC Positioning using DNN via WkNN	DNN, WkNN	93	0.43	0.30	Explored DNN with Weighted KNN, showing improved performance	[36]
3	ANN-based Regression for VLP	ANN	90	0.52	0.35	Demonstrated effectiveness of ANN for Regression in VLP	[13]
4	Neural Network-based Accurate VLP	NN	95	0.36	0.28	Utilized neural networks, emphasizing positioning accuracy	[43]
5	Modified Path-loss Model for Indoor VLP	Modified Path-loss	91	0.45	0.30	Proposed modified path-loss model with sparse fingerprints for indoor precision	[2]
6	Tilted-angle Insensitive RSS VLP using DNN	DNN	94	0.40	0.29	Developed DNN model with synthetic data for angle-insensitive RSS, improving robustness	[59]
7	Comprehensive ML Approach for VLP	ML	87-94	0.35-0.55	0.25-0.4	Investigated various ML techniques for enhancing VLP systems	[42]
8	Multiple-Photodiode Receiver with ML for VLP	ML	96	0.33	0.23	Used multiple-photodiode receivers and ML for high accuracy in VLP	[2]

3.3.6.6 Advantages and Disadvantages of KNN

K-Nearest Neighbors (KNN) offers several notable benefits as well as some drawbacks that must be considered when applying the algorithm to different problems:

Advantages of KNN in VLP:

- **Simplicity and Intuitive Understanding:**

  One of the significant strengths of K-Nearest Neighbors (KNN) in Visible Light Positioning (VLP) is its simplicity. The algorithm is straightforward to understand and implement in VLP systems without having an elaborate training process. This is particularly beneficial in VLP applications where real-time deployment is critical, as KNN performs all computations during the positioning phase, making it suitable for low-latency environments.

- **Flexibility Across Different Distance Metrics:**

  KNN's versatility is highly advantageous for VLP, as it can be adapted to different distance metrics

depending on the signal characteristics and positioning environment. For instance, Euclidean distance may be used for straightforward signal strength measurements, while Cosine distance could be applied in environments where the angle of light incidence affects signal perception. This flexibility enables KNN to be customised for various VLP setups, enhancing its adaptability to different signal variations and environmental conditions [55], [36].

- **Non-parametric Nature:**

  KNN is a non-parametric algorithm, meaning it does not rely on assumptions about the underlying distribution of light signals or environmental factors. In VLP, this is especially useful as the indoor environment and signal propagation can be unpredictable, with irregularities like light reflections and obstacles. KNN's ability to handle such irregular data patterns makes it well-suited to VLP, where the environment often defies idealised assumptions [13], [42].

Disadvantages of KNN in VLP:

- **Computational Complexity:**

  While KNN's simplicity is beneficial, its computational complexity can pose challenges in VLP systems, especially with large datasets or dense reference fingerprint maps. KNN must calculate the distance from the query point to every reference point in the dataset, which becomes computationally expensive as the size of the reference dataset increases. In real-time VLP, this can delay determining the user's position, especially in large indoor spaces or when frequent recalculations are needed to maintain accuracy [2], [55].

- **Memory Inefficiency:**

  KNN's memory inefficiency can also be a drawback in VLP, as the algorithm requires storing the entire reference fingerprint map. In indoor positioning systems where signals and light fingerprints must be recorded across multiple locations, this can result in high memory usage. This issue becomes more pronounced as the number of reference points and signal features (e.g., light intensity, angles) increases, making KNN less feasible for systems with limited storage capacity [55].

- **Sensitivity to High Dimensionality:**

  In VLP, signal fingerprints may involve multiple features such as light intensity, angle of incidence, and the distance between the light source and receiver, leading to high-dimensional data. KNN suffers from the curse of dimensionality, where the distinction between nearby and distant signal fingerprints becomes less clear as the number of features increases. This can degrade the algorithm's performance, making it harder to differentiate between closely spaced positions and decreasing positioning accuracy in environments with complex signal interactions [42], [36].

- **Imbalance in Signal Fingerprints:**

  KNN can struggle with imbalanced signal datasets in VLP, where certain areas of an indoor environment may have more densely collected signal data than others. For instance, regions close to light sources might have more reference fingerprints, while those in shadowed or less illuminated areas have fewer. In such cases, KNN may disproportionately favour the well-represented regions, leading to biased positioning results. Techniques such as weighted KNN or oversampling methods may be required to counteract this imbalance and ensure more accurate positioning across the entire environment [2], [55].

3.3.6.7 Variations and Improvements

Enhancements and hybrid approaches can be applied to overcome the limitations of the KNN algorithm. These include weighted KNN, which assigns different weights to neighbours based on their distance, giving more importance to closer data points. Optimised search algorithms such as KD-trees and Ball Trees significantly reduce search time in large datasets, improving KNN's efficiency. These modifications address KNN's performance issues and expand its applicability to more complex scenarios. Some such changes are discussed in greater detail in the sections that follow.

3.4 K-Nearest Neighbors (KNN) with Signal Strength Clustering

KNN with Signal Strength Clustering is a method that enhances the accuracy of position estimation by leveraging both the simplicity of KNN and the organisation of signal strength data through clustering techniques, such as k-means [61].

1. **K-Nearest Neighbors (KNN):** KNN estimates the position of the receiver/device by comparing it to known reference points in a given feature space. Each reference point has a known position and an associated RSS value. KNN calculates the distance between the receiver/device and every reference point to determine its location. The algorithm then selects the K nearest reference points and uses their positions to estimate the receiver/device's location [50].

2. **Signal Strength Clustering:** Before applying KNN, signal strength data is often clustered using algorithms like k-means. Clustering organises the data into smaller groups based on similarities in RSS values. This process reduces the complexity of the data, making it easier to manage and process. By grouping similar signal strengths, clustering highlights key signal characteristics and simplifies the representation of the environment, leading to a more accurate and efficient KNN application [30].

The process of WKNN clustering involves the following steps:

1. **Data Collection:** RSS data is collected from multiple access points or sensors across the environment.

2. **Clustering:** The collected RSS data is clustered or organised using k-means or similar algorithms. Each cluster represents a group of RSS values with similar characteristics [30].

3. **Weighted KNN (WKNN):** Once clustering is done, WKNN assigns weights to each cluster based on its importance in estimating the receiver/device position. Weights are based on factors like cluster density or the variance in signal strength within the cluster. Clusters with higher significance in predicting the receiver/device position receive higher weights [19].

4. **Position Estimation:** Finally, WKNN uses the weighted clusters to estimate the position of the receiver/device. The K nearest clusters are selected based on their weights, and their positions are used to compute the estimated position.

Mathematically, WKNN can be represented as:

$$\hat{Y} = \frac{\sum_{i=1}^{K} w_i \cdot Y_i}{\sum_{i=1}^{K} w_i}$$

Where:

- $\hat{Y}$ is the estimated position of the target point.
- Y_i is the position of the i-th nearest cluster.
- w_i is the weight assigned to the i-th nearest cluster.

The diagram 3-4 below describes the process of KNN with signal strength clustering.

The contributions of KNN with Signal Strength Clustering in VLP systems can be summarised as follows:

- Enhanced Indoor Localisation: Clustering algorithms like k-means create precise maps. KNN-based indoor localisation systems achieve higher accuracy, particularly in Wi-Fi-based setups. The clustering of signal strength data simplifies the complexities of indoor environments, leading to more reliable positioning [30].
- Efficient Network Management: Signal strength clustering can significantly benefit mobile wireless sensor networks beyond localisation. Clustering nodes based on received signal strength enables efficient network management, improved energy conservation, and streamlined data processing [51].
- Adaptive Positioning Precision: Combining weighted KNN algorithms with advanced clustering techniques, such as affinity propagation clustering, leads to adaptive positioning methods. These methods dynamically select the most relevant reference points, improving precision [12].

3.4.1 Frequency Modulation in Visible Light Positioning Systems

When selecting a frequency modulation technique to enhance Visible Light Positioning (VLP) systems, the choice largely depends on the specific requirements of the VLP application. Accuracy, data rate, complexity, and environmental conditions should be considered. Based on recent studies, several frequency modulation techniques have been suggested for VLP systems.

- **Binary Frequency-Shift Keying (BFSK) Modulation:**According to a study conducted by Beguni et al. (2023) [11], BFSK has demonstrated improved noise resilience. This makes it a good choice for

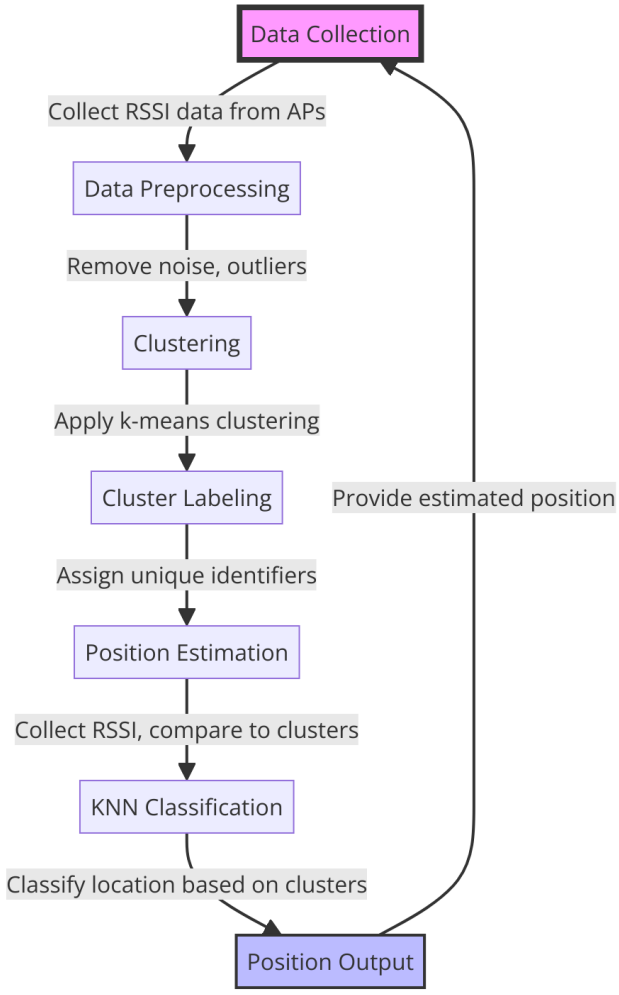


Figure 3-4: Diagram illustrating the K-Nearest Neighbors (KNN) with Signal Strength Clustering in VLP process

VLP systems that operate in noisy or interference-prone environments. BFSK is more suitable for applications where the ability to withstand noise is more important than achieving high data rates.

- **Orthogonal Frequency Division Multiplexing (OFDM):** OFDM (Orthogonal Frequency Division Multiplexing) has different variants like DCO-OFDM and Flip-OFDM, which are suitable for high data rate requirements. OFDM is often used in VLP (Visible Light Positioning) systems, where data communication efficiency is crucial. Saengudomlert and Buddhacharya (2022) [44] have discussed its usefulness, particularly in scenarios requiring simultaneous high-speed data transmission and accurate positioning.

- **Orthogonal Time Frequency Space (OTFS) Modulation:** OTFS may benefit VLP systems in dynamic environments by offering robustness to channel variations, which could be particularly relevant for mobile positioning applications within VLP systems, as highlighted in the work of Saengudomlert et al. (2022) [10].

- **Index Modulation in OFDM (CRIP-OFDM):** This modulation technique benefits VLP systems requiring efficient data transmission with controlled lighting conditions. It balances data transmission and environmental factors such as brightness control, as discussed at the Asia Communications and Photonics Conference and Exhibition [9].

- **Pulse Amplitude Modulation (PAM):** DFT Spread-Optical PAM, also known as PAM, is a promising option for VLP (Visible Light Communication) systems that require high spectral efficiency. This technology is particularly suitable for crowded indoor environments with high traffic, where efficient spectrum utilisation is essential. Saied et al. (2022) [45] have presented the benefits of DFT Spread-Optical PAM in their research.

The optimal choice for a VLP (Visible Light Communication) system relies on the specific demands and limitations of the application, which may include factors like data transmission needs, environmental conditions and system complexity. Each technique has its unique benefits and drawbacks, which must be assessed carefully to ensure the best possible performance of the VLP system.

4. RESEARCH METHODOLOGY

This chapter outlines the research methodology for implementing the Visible Light Positioning (VLP) system, focusing on indoor localization utilizing multiple photodiodes and fingerprinting. The research systematically develops a theoretical framework through a comprehensive literature review, guiding the algorithm selection and design for the VLP system. This methodology combines theoretical modelling with practical simulations to evaluate the proposed system's performance.

4.1 Research Design

The research design was developed through an iterative process based on concepts discussed in previous Chapters. This framework draws from the investigation of VLP systems and KNN with signal strength clustering, frequency modulation techniques, and the addition of ROI selection, allowing for a comprehensive design.

4.2 Simulation Environment

MATLAB is used as the simulation environment due to its robustness and the availability of extensive toolkits for signal processing and machine learning. The simulations are designed to emulate ideal light sources and controlled indoor environments.

4.3 Algorithm Development

The Algorithm for the proposed system is split into two main phases. The main focus is to create and enhance the fingerprint-based VLP algorithm to optimize indoor positioning accuracy using multiple photodiodes and ROI selection. Key factors considered during algorithm development included:

- Signal Processing: How to efficiently process and interpret the signals the photodiodes receive.

- Fingerprinting Technique: Creating a database of known signal patterns at various locations within the indoor environment that would allow for maximum accuracy.

- Localization Algorithms: Weighted k-nearest Neighbors (WkNN) and Weighted K-nearest neighbours with signal clustering approaches were employed to match real-time readings with the fingerprint database.

Figure 4-1 illustrates the two phases of the proposed system, the online and offline phases, which are discussed in greater detail below.

4.3.0.1 Offline Phase

During the offline phase, on the left of 4-1, the room's LED layout and the multi-photodiodes' behaviour are simulated and modelled, creating our base environment for VLP. The process begins by defining the key simulation parameters, which include the room size, the number of LEDs required to meet lux requirements, and the lumens per LED. Once these parameters are set, an optimal LED layout is determined to fit the room. This layout is crucial for providing sufficient lighting coverage and positioning accuracy. Alternate LEDs are selected, and unique data IDs are assigned to identify smaller areas or quadrants within the room. This methodology is crucial for simulating and modelling the room's LED layout and multi-photodiodes' behaviour.

The fingerprint database is also constructed in the offline phase. It captures the correlation between RSS values and frequency IDs at different reference locations in a room. This is achieved through the following steps: The room area is dynamically sampled based on room size, generating sample points to measure RSS and VLC IDs. The Lambertian emission model determines the RSS value at each sample point in the room. The impact of ambient light on RSS values is accounted for by adding a noise component. We introduce a noise term N_{ambient}, where N_{ambient} follows a Gaussian distribution with mean μ and variance σ^2:

$$N_{\text{ambient}} \sim \mathcal{N}(\mu, \sigma^2)$$

The total received signal strength P_{total} is the sum of the LED's RSS and the ambient light noise.

$$P_{\text{total}} = P_{\text{LED}} + N_{\text{ambient}}$$

Substituting the Lambertian expression for P_{LED}, and μ, the mean of the Gaussian distribution representing the average ambient light. Taking into account the Uncertainty introduced by the ambient light given by the variance of Gaussian noise [17]. The complete Lambertian emission model equation, accounting for Gaussian-distributed ambient light, is:

$$P_{\text{total}} = \frac{P_t \cdot m \cdot \cos^m(\theta)}{d^2} \cdot A \cdot T(\psi) \cdot g(\psi) \cdot \cos(\psi) + N(\mu, \sigma^2)$$

Sampling, RSS calculation, and ambient light correction are combined to build the fingerprint database for VLP.

4.3.1 Online Phase

During the online phase, on the right of 4-1, the multi-photodiode receiver is virtually moved to different positions within the room to collect RSS/VLC data. The Region of Interest (ROI) is identified by detecting the highest frequency ID in the collected data. The fingerprint database is then narrowed down to a smaller area where the corresponding LED ID is located. The receiver's position is then estimated using the Weighted k-nearest Neighbors (WKNN) algorithm and signal clustering.

4.4 Performance Testing and Benchmarking

To assess the performance of the developed VLP system, a series of tests and benchmarks were conducted, which included: Accuracy Assessment: The system's precision in determining the correct position was rigorously tested under various conditions. Comparative Analysis: The system's performance was compared with existing methods to evaluate improvements or advancements. Scalability and Robustness Tests: The system was tested for its effectiveness in environments of different sizes and under various ambient light

conditions.

4.5 Data Analysis

The data collected from the simulations were subjected to comprehensive analysis using MATLAB statistical tools to extract meaningful insights regarding the performance of the Visible Light Positioning (VLP) system. The analysis focused on assessing the system's accuracy, reliability, and efficiency to provide valuable insights into its strengths and limitations. The following key metrics were employed to evaluate the performance of the VLP system as they are commonly used in VLP systems to asses the accuracy between the estimated positions given by the VLP system when compared to the actual position:

- **Mean Absolute Error (MAE)**: The Mean Absolute Error (MAE) is a metric that calculates the average absolute difference between the estimated position and the actual position as given in 4.1. It considers the magnitude of errors in the position estimation process irrespective of direction. The lower the MAE value, the higher the accuracy of the position estimation.

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|P_{i,\text{estimated}} - P_{i,\text{true}}\right| \tag{4.1}$$

  where $P_{i,\text{estimated}}$ is the estimated position, $P_{i,\text{true}}$ is the true position, and N is the number of position estimates.

- **Root Mean Square Error (RMSE)**: The Root Mean Square Error (RMSE) is a measure of the accuracy of a position estimation. It is calculated by taking the square root of the average of the squared differences between the estimated and actual positions as shown in 4.2. RMSE is a valuable metric as it penalizes significant errors more heavily than Mean Absolute Error (MAE). The lower the RMSE value, the more accurate the estimation.

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(P_{i,\text{estimated}} - P_{i,\text{true}}\right)^2} \tag{4.2}$$

  where $P_{i,\text{estimated}}$ is the estimated position, $P_{i,\text{true}}$ is the true position, and N is the number of position estimates.

- **95th Percentile Error**: The 95th percentile error is a measure that shows the position error for which 95% of the estimated positions fall below. It is calculated using equation 4.3. This measure can help

to understand the distribution of errors and identify outliers or cases with exceptionally high errors. A lower 95th percentile error indicates better performance and higher consistency in position estimation.

$$P = 0.95 \times (N + 1) \tag{4.3}$$

where N is the total number of data points. If P is not an integer, interpolate between the nearest ranks.

These metrics collectively provide a comprehensive assessment of the accuracy of the position estimation process in the VLP system. The system's effectiveness and reliability can be evaluated by analysing these metrics, enabling informed decisions regarding system optimization and improvements.

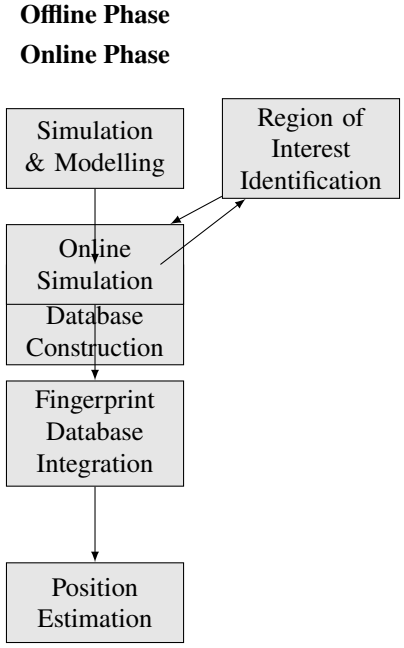


Figure 4-1: Proposed ROI & FingerPrintbased VLP Algorithm

4.6 LED Selection and Region Identification

In the indoor positioning system, specific Light Emitting Diodes (LEDs) strategically placed within the environment are utilized for region selection purposes. These LEDs are carefully chosen and modulated with different frequencies within a predefined range to generate unique signal patterns or IDs. Frequent modulation enables the system to distinguish between signals emitted by different LEDs, thereby enhancing the accuracy

of the positioning system.

4.6.1 Frequency Modulation for Signal Differentiation

The frequency modulation process involves varying the frequency of the carrier signal according to the amplitude of the message signal. Mathematically, this can be expressed as:

$$s(t) = A_c \cdot \sin(2\pi f_c t + \phi) \cdot \sin(2\pi f_m t) \tag{4.4}$$

Where:

- $s(t)$ represents the modulated signal,
- A_c is the carrier signal amplitude,
- f_c is the carrier signal frequency,
- f_m is the message signal frequency, and
- ϕ is the phase angle.

The flexibility in LED selection allows for the modulation of specific LEDs with varying frequencies, enabling more precise identification and localization of regions within the indoor environment. For instance, the function now incorporates a selection pattern parameter, such as assigning a selection pattern of two—every second LED is chosen to have an ID, providing greater flexibility.

4.6.2 Dynamic ID Assignment

In this updated version, IDs are assigned based on a range of frequencies, simulating a realistic scenario where each LED may be modulated at a different frequency. The range of frequencies, customizable via the parameter *idFrequencyRange*, is chosen to be 800 Hz and higher, ensuring that the modulation frequencies are invisible to the human eye.

4.6.3 Improved Area Identification

Rather than manually dividing the space into quadrants based on the median, this system version employs k-means clustering to dynamically identify areas based on the spatial distribution of the selected LEDs. This approach enhances adaptability to various room shapes and configurations of LEDs, thereby improving the accuracy and robustness of region identification.

Furthermore, the LED assignment function returns the IDs and positions of the selected LEDs, providing comprehensive information for subsequent localization algorithms.

4.6.4 Pulse Amplitude Modulation (PAM) Modulation

Pulse Amplitude Modulation (PAM) is adopted as the modulation scheme for the VLP system. PAM is a promising option for Visible Light Communication (VLC) systems requiring high spectral efficiency. This technology is particularly suitable for crowded indoor environments with high traffic, where efficient spectrum utilization is essential. Saied et al. (2022) [45] have presented the benefits of Discrete Fourier Transform (DFT) Spread-Optical PAM in their research.

The PAM modulation scheme involves modulating the amplitude of the carrier signal based on the message signal. Mathematically, PAM can be represented as:

$$s(t) = A_c \cdot m(t) \cdot \sin(2\pi f_c t + \phi) \tag{4.5}$$

where: m(t) is the message signal. A flow diagram illustrating the LED ID assignment process is depicted in Figure 4-2. This diagram outlines the sequential steps in selecting and assigning LED IDs based on the



Figure 4-2: Flow Diagram for LED ID Assignment Process

specified parameters and modulation scheme. This pseudo-code outlines the simulation process for a VLP system, including room setup, fingerprint database generation, parameter definition, position estimation, and

region of interest selection.

Algorithm 1 Simulation of VLP System

Initialize room and LED parameters
$roomLength \leftarrow 10$ ▷ Room length in meters
$roomWidth \leftarrow 10$ ▷ Room width in meters
$ledPower \leftarrow 0.5$ ▷ LED power in watts
$m \leftarrow -1 \times \log_2(\cos(\theta))$ ▷ Lambertian order
$ambientLight \leftarrow 0.1$ ▷ Ambient light level in watts
$samplingDensity \leftarrow 2$ ▷ Samples per meter
$luxRequirement \leftarrow 50$ ▷ Lux requirement
$lumensPerLED \leftarrow 200$ ▷ Lumens per LED
$FOV \leftarrow 20$ ▷ Field of view in degrees
Initialize LED grid
$numPositions \leftarrow 20$ ▷ Number of random positions
$K \leftarrow 10$ ▷ Number of nearest neighbors
$distanceThreshold \leftarrow 5$ ▷ Threshold for match
$numClusters \leftarrow 3$ ▷ Number of clusters
$distanceMetric \leftarrow' manhattan'$ ▷ Distance metric
Create an LED grid
Generate random receiver positions
Select LEDs for modulation and assign IDs
Build the fingerprint database
Simulate receiver positions and observations
Filter fingerprint database for ROI
Estimate receiver positions
Plot real vs estimated positions
Calculate positioning performance
Create error visualisations
Display performance metrics

4.6.5 Region of Interest Identification

In the online phase of the VLP system, the Region of Interest (ROI) is determined based on the received VLC data. This process involves frequency analysis of the data to identify the LED sources closest to the receiver. The dominant VLC ID, corresponding to the highest frequency components in the received signal, selects the region.

4.6.6 Frequency Analysis

The Fast Fourier Transform (FFT) is used to improve ROI selection. Analyzing the frequency components of the Visible Light Communication (VLC) signal is more effective using this technique. Assuming the VLC data consists of a time series of signal strengths, we can derive frequency components that indicate proximity

to LED sources. Consider a 4x4 room with two standard and frequency-modulated LEDs, as shown in Figure 4-3. A receiver is positioned at a random location within the room. The closest LEDs can be identified by analyzing the received signal strengths and applying frequency analysis, and the corresponding region can be assigned to the receiver's position.

- **Collect Time Series Data:** Assume *vlcData* includes time series data of signal strengths for each LED, allowing us to apply FFT and analyze frequency components.
- **Apply FFT:** For each LED, apply FFT to the time series data to identify dominant frequency components.
- **Identify Proximity Based on Frequency Components:** Use the frequency component magnitudes to determine which LED the receiver is likely closest to, assuming that certain frequency patterns or strengths indicate proximity.

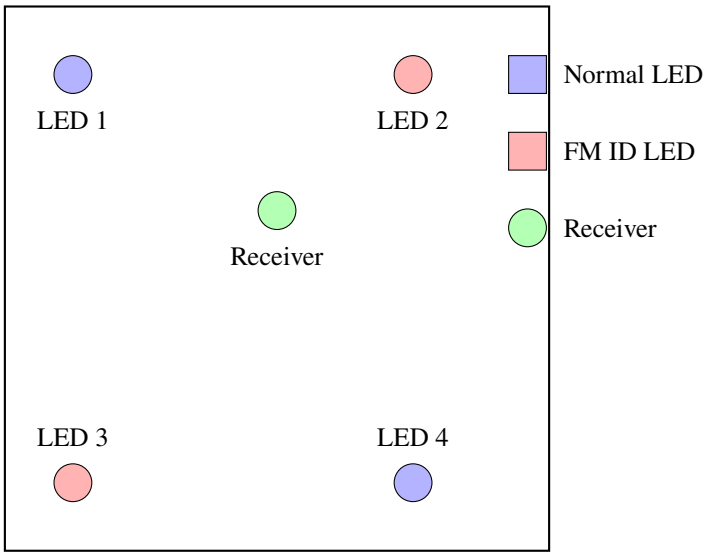


Figure 4-3: Example of LED Selection and Room Layout

In each ROI, the signals from the frequency-modulated LEDs are analyzed for parameters such as signal strength and their unique IDs.

4.6.7 Preprocessing Algorithm

A preprocessing algorithm is developed to process the signal data from the selected frequency-modulated LEDs. This algorithm filters out noise and irrelevant data, ensuring that only pertinent signal information and the unique IDs of frequency-modulated LEDs are used for ROI identification.

4.6.8 Integration with Fingerprinting Database

The signal data preprocessed from each ROI is combined into the fingerprinting database. This combination makes it possible to match real-time signal readings with the database more efficiently and accurately during the localization process. This is achieved by reducing the size of the fingerprint database that needs to be matched.

4.6.9 Simulation of Region Selection Preprocessing

The effectiveness of region selection preprocessing can be tested using MATLAB simulations. It is essential to evaluate how using frequency-modulated LEDs impacts the accuracy of indoor localization. The efficiency of the region selection algorithm can also be determined in various indoor settings. Additionally, assessing how or when preprocessing improves localization precision and reduces computational load is essential.

4.6.10 Performance Metrics for Preprocessing

The performance of the region selection preprocessing is quantified using specific metrics, including: Localization Accuracy: The system's precision in correctly identifying the user's location within a particular ROI. Computational Efficiency: The reduction in computational load and processing time due to the preprocessing step. System Robustness is the ability of the system to maintain accuracy under different environmental conditions, such as varying light levels and obstacles.

4.7 Determining Parameters in the WKNN Algorithm with Signal Clustering

The parameters in the WKNN algorithm, including the value of K and the choice of distance metric (Manhattan), were determined systematically to optimize the algorithm's performance. The following steps outline the methodology for determining these parameters and the WKNN with signal clustering functions or algorithms.

4.7.1 Parameter Tuning for WKNN Algorithm with Signal Clustering

Algorithm 2 Parameter Tuning for WKNN Algorithm with Signal Clustering

```
Define a range of parameter values
K_values ← [1, 3, 5, 7, 9]                          ▷ Example range of K values
distance_metrics ← {'euclidean','manhattan','minkowski'}   ▷ Example distance metrics
Initialize variables to store performance metrics
best_accuracy ← 0
optimal_K ← −1
optimal_metric ←''
Determine number of clusters, numClusters, using an cross-validations with WKNN method
for K in K_values do
    for metric in distance_metrics do
        Train WKNN model with K neighbors and metric distance metric with signal clustering
        Evaluate the performance of the WKNN model on the validation set
        if performance > best_performance then
            best_K = K
            best_metric = metric
            best_performance = performance
            if average_accuracy > best_accuracy then
                best_accuracy ← average_accuracy
                optimal_K ← K
                optimal_metric ← metric
            end if
        end if
    end for
end for
Display optimal parameters
print "Optimal K value:" optimal_K
print "Optimal distance metric:" optimal_metric
```

Algorithm 3 Estimate Positions with WKNN and Clustering

```
function estimatePositionsWithWKNNAndClustering(fingerprintDatabase, observedRSS, K, numClusters)
    Input: Fingerprint database, observed RSS signals, K, numClusters
    Output: Estimated positions
    Perform k-means clustering on observed RSS signals
    clusterIndices = clusterRSSSignals(observedRSS, numClusters)
    numObservations = size(observedRSS, 1)
    estimatedPositions = zeros(numObservations, 2)
    for obsIndex = 1 : numObservations do
        currentClusterIndex = clusterIndices(obsIndex)
        clusterFingerprintDatabase = filterFingerprintDatabase(fingerprintDatabase, currentClusterIndex)
        if isempty(clusterFingerprintDatabase) then
            continue                    ▷ Skip if no matching fingerprints for the current cluster
        end if
        observedRSSVector = observedRSS(obsIndex, :)
        estimatedPositions(obsIndex, :) = applyWKNN(clusterFingerprintDatabase, observedRSSVector, K)
    end for
    return estimatedPositions
end function
```

Algorithm 4 Cluster RSS Signals

```
function clusterRSSSignals(observedRSS, numClusters)
    Input: Observed RSS signals, numClusters
    Output: Cluster indices
    Clusters the observed RSS signals into numClusters clusters based on similarity.
    [clusterIndices, ] = kmeans(observedRSS, numClusters,'Distance','sqeuclidean','MaxIter', 1000)
    return clusterIndices
end function
```

Algorithm 5 Apply WKNN

```
function applyWKNN(fingerprintDatabase, observedRSS, K)
    Input: Fingerprint database, observed RSS signal, K
    Output: Estimated position
    RSSValues = fingerprintDatabase(:, 3 : end)
    positions = fingerprintDatabase(:, 1 : 2)
    distances = zeros(size(RSSValues, 1), 1)
    for i = 1 : size(RSSValues, 1) do
        fingerprintRSS = RSSValues(i, :)
        distances(i) = relevant distance metric calculation
    end for
    [ , sortedIndices] = sort(distances)
    kNearestIndices = sortedIndices(1 : K)
    return estimatedPosition
end function
```

4.7.2 Estimate Positions with WKNN and Clustering

Systemically determining parameters in the WKNN algorithm with signal clustering ensures optimal performance in indoor positioning and classification tasks.

4.8 Methodology Summary

This methodology ensures the systematic development, evaluation, and optimization of the Multi-Photodiode Fingerprint-Based Visible Light Positioning (VLP) system, providing a robust framework to address the research questions and objectives outlined in the study. By integrating theoretical modelling with practical simulations, the methodology facilitates a comprehensive exploration of indoor localization techniques, enabling in-depth investigation into the system's performance under various conditions. The system's accuracy, reliability, and efficiency are assessed through performance testing, allowing the research objectives of enhancing indoor positioning precision to be pursued. Moreover, by leveraging advanced techniques such as frequency modulation, dynamic ID assignment, and signal clustering, the methodology improves the system's adaptability and precision, thus aligning with the larger goal of contributing to the advancement of indoor localization technologies to meet the demands of modern navigation systems.

5. RESULTS AND ANALYSIS

5.1 Computational Complexity Analysis

Assessing the computational complexity of a system involving room setup for Visible Light Positioning (VLP), generating a fingerprint database, simulating real-time positions, and estimating positions using K-Nearest Neighbors (KNN) involves analysing each system component. Table 5.1 divides the complexity into critical parts of the system. The variables within the system represent the following measurements:

- N: Total number of reference points or LEDs in the system.
- M: Number of measurements or data points collected per reference point.
- P: Total number of positions simulated in real-time scenarios.
- Q: Number of queries or location estimations made using the system.
- C: Number of clusters used in the KNN with signal strength clustering.

Table 5.1: System Components and Their Complexities

System Component	**Complexity**
Room Setup with LEDs	$O(1)$
Generating Fingerprint Database	$O(N \times M)$
Simulating Real-Time Position	$O(P \times M)$
Estimating Positions Using KNN	$O(Q \times N)$
KNN with Signal Strength Clustering	$O(Q \times (N + C))$

Computational complexity is generally determined by the system's order or the system's nth of the highest order in terms of complexity. Depending on the specific implementations and optimisations applied, this is likely the estimation step due to its potential $O(Q \times N)$ or $O(Q \times (N + C))$ complexity. However, the fingerprint database generation is also significant in particular scenarios ($O(N \times M)$ complexity), especially for large rooms with many LEDs and dense sampling is required to build the fingerprint database.

Room Setup complexity is constant across room sizes, confirming the $O(1)$ notation, indicating that the setup process does not scale with the room size. In contrast, both Fingerprint Database Generation and Real-Time Positioning display linear complexity growth as the room size increases, denoted by $O(N \times M)$

and $O(P \times M)$ respectively. These relationships suggest a direct proportionality between the number of measurements and the room's scale.

The complexity for Estimating Positions using KNN has a linear increase in figure **??**, albeit at a lower rate, consistent with its $O(Q \times N)$ complexity. This could be due to a constant number of nearest neighbours considered (N) or a linear relationship between the number of data points (Q) and room size. KNN with Signal StKNN with Signal Strength Clustering initially performs similarly; however, as the room size increases, we have a much steeper increase in complexity, in line with the $O(Q \times (N + C))$ expression provided above. Demand for KNN with signal clustering depends very much on the size of the data set.

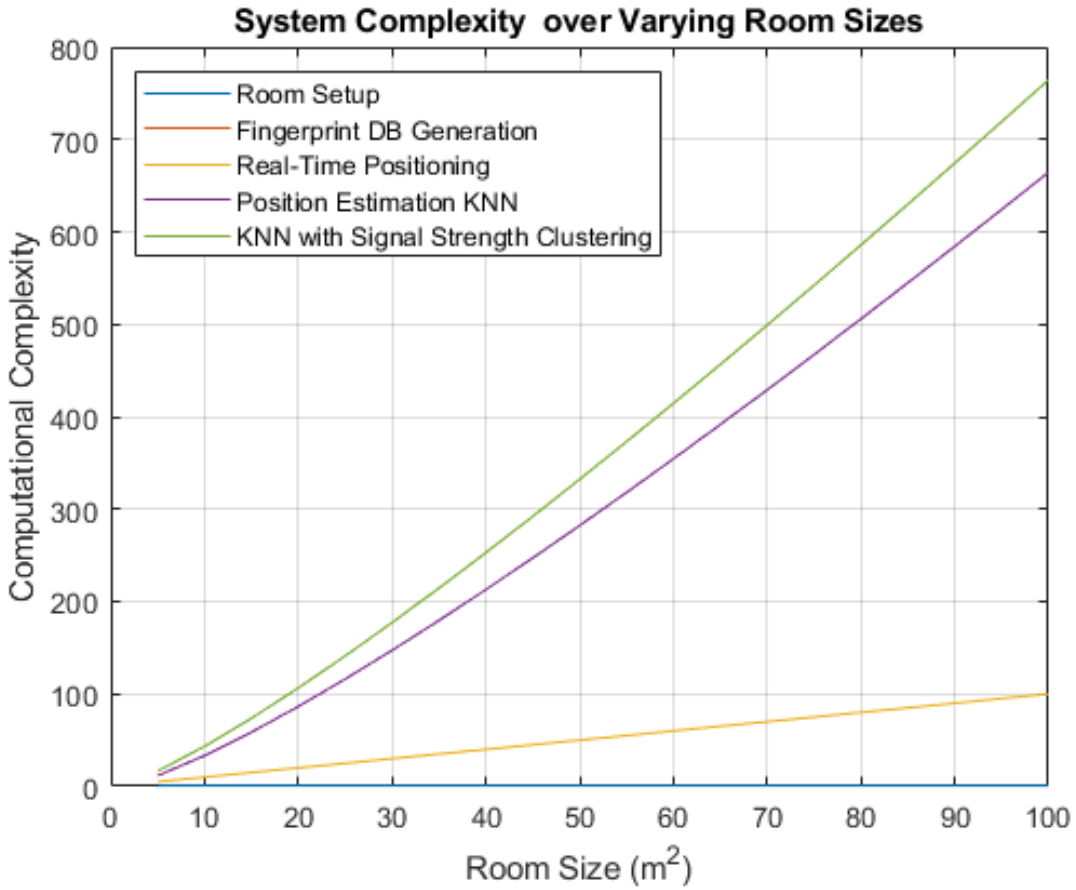


Figure 5-1: System Complexity over Varying Room Sizes

Figure **??** highlights the increasing demands on data processing and estimation accuracy as the room or data set size increases. The components mentioned above are the proposed system's primary consumers of computational resources. This suggests that each algorithm's component will likely require optimisation and

possibly parallel processing to manage the computational load in large-scale or real-life VLP applications with much larger room sizes. Optimisation scalability, particularly for data-intensive processes such as database generation and clustering-based estimation methods of the proposed VLP system if it were to be applied to larger data sets or implemented in rea-life.

5.2 Results of VLP system with WKNN Without ROI Selection & Room size 4mx4m

Analysis of figure 5-2 reveals several noteworthy observations. The system's performance showed considerable variability across different iterations within the small room, with a Mean Absolute Error (MAE) of 4.015 meters, indicating apparent inaccuracies. Despite most errors falling below this threshold, there is a significant tail of larger errors with a 95th percentile error of 6.208 meters. The errors were notably reduced by optimising the parameters of the WKNN algorithm, as detailed in 4, and implementing the ROI selection algorithm. Additionally, adjustments in receiver parameters and environmental factors may contribute to further improvements.

Table 5.2: Average Performance Metrics of The System Without ROI & Room Size 4mx4m

Metric	**Value (meters)**
Mean Absolute Error (MAE)	4.015
95th Percentile Error	6.208

5.3 Results of VLP system with WKNN with Signal Clustering Optimization and ROI Selection & Room Size 4mx4m

The VLP (Visible Light Positioning) system was improved through an optimisation process that involved optimised selection of the parameters for the Weighted K-Nearest Neighbors (WKNN) algorithm, implementing signal filtering techniques and fine-tuning the Region of Interest (ROI) selection based on frequency. As a result of this process, the WKNN algorithm's parameter values were refined, significantly improving position estimation accuracy.

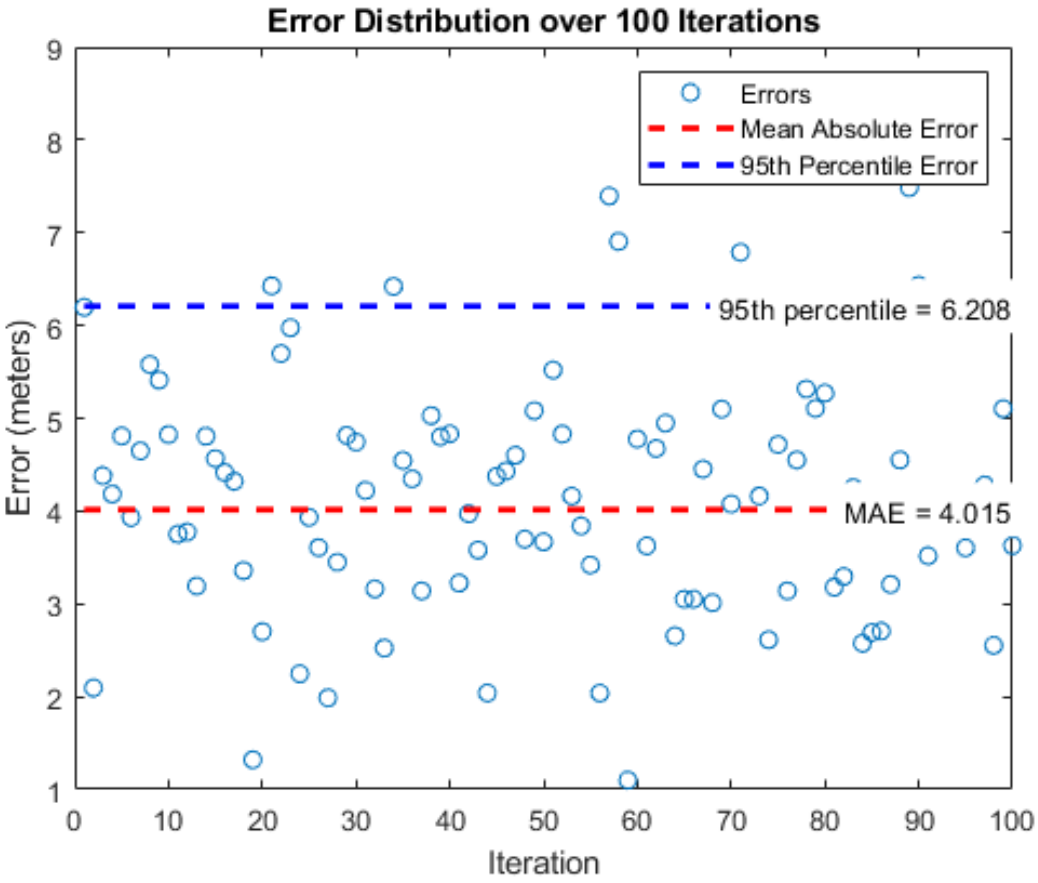


Figure 5-2: Performance Metrics of VLP System Without ROI & Room Size 4mx4m

The performance metrics of the optimised VLP system, including WKNN and ROI selection, are presented below:

Table 5.3: Average Performance Metrics of Optimized VLP System & Room Size 4mx4m

Metric	**Value (meters)**
Mean Absolute Error (MAE)	3.8
Root Mean Square Error (RMSE)	4.1
95th Percentile Error	5.5

The error distribution plot 5-3 illustrates enhanced accuracy compared to its non-optimised counterpart, boasting a Mean Absolute Error (MAE) of 3.800 meters and with 95% of errors falling below 5.500 meters. Although notable improvements are observed, occasional errors of up to 7 meters indicate noticeable deviations from actual positions. While integrating signal clustering and ROI selection techniques has effectively constrained error distribution, suggesting a more dependable system, further optimisation is required to

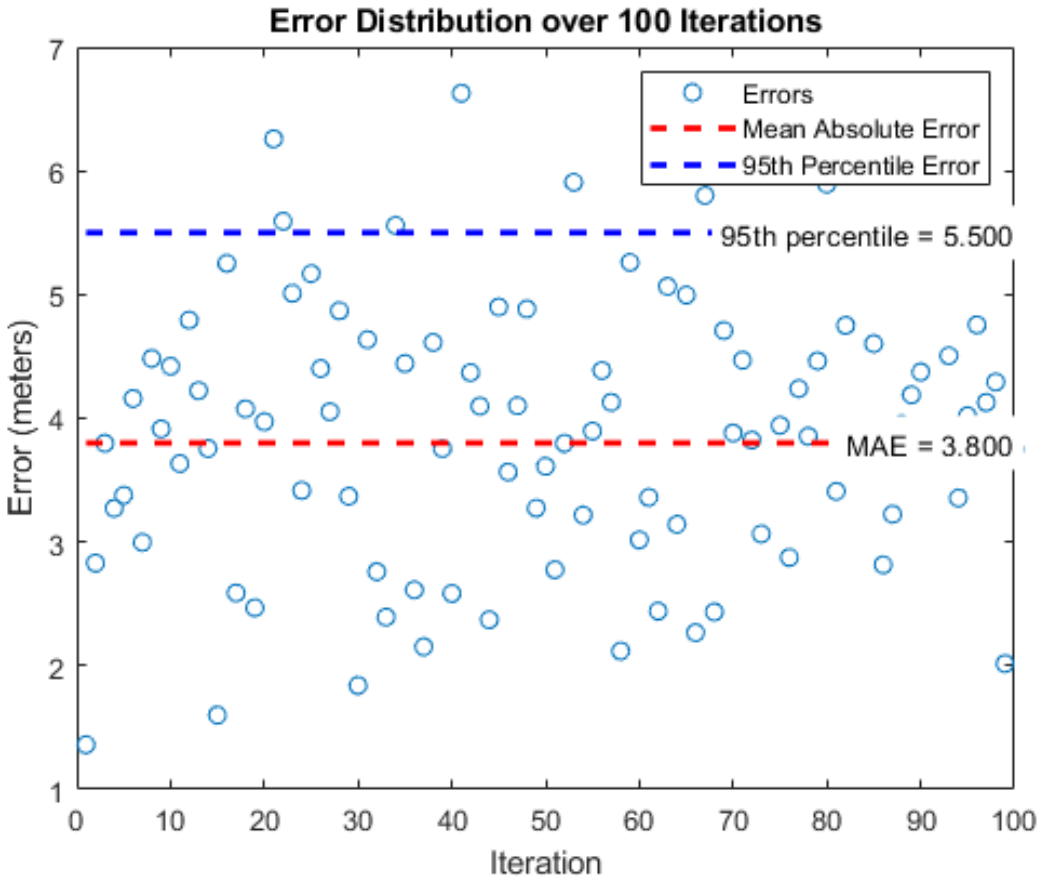


Figure 5-3: Performance Metrics of Optimized VLP System & Room Size 4mx4m

ensure reliable data and precision in practical deployment scenarios. Applications.

Table 5.4: Errors of Optimized VLP System & Room Size 4mx4m under Higher Ambient Light Conditions

Ambient Light Level (watts)	**Error (meters)**
0.1	2.5
0.2	4.0
0.3	6.0
0.4	8.5
0.5	11.5

The RSS calculation, given by the Lambertian emission model described in **??**, incorporates the ambient light level alongside the LED power and distance-dependent Lambertian emission. As ambient light increases, the additive component in the RSS formula becomes more significant, potentially overshadowing the LED signal. Consequently, the system may experience inflated errors in position estimation, particularly in environments with intense ambient illumination. Table 5.4 demonstrates the optimised VLP system's sensitivity

to changes in ambient light within a 4-meter by 4-meter room. As the ambient light intensity increases from 0.1 to 0.5 watts, the estimated positioning error of the system also increases from 2.5 to 11.5 meters. This suggests that there may be an exponential relationship between ambient light levels and positioning error, particularly at higher ambient light intensities. The system's susceptibility to changes in lighting conditions highlights the need for solid signal processing techniques, strict environmental controls, or hybrid positioning methods to improve accuracy in more realistic simulations with diverse indoor lighting scenarios or real-world applications of the system. To improve system performance in areas with higher ambient light, several strategies can be explored:

- Improved LED Model: Changing the LED model such that the LEDs have a higher Luminous flux (brightness) will increase the SNR, thus helping to differentiate the LED's signal from ambient light.
- Enhanced RSS Simulation: Developing an RSS simulation model that accounts for multipath propagation and reflection offers a more comprehensive representation of real-world signal propagation. This would account for wall reflections, diffraction, and interference patterns to more accurately simulate the received signal strength and be a more accurate representation of how the system would perform if implemented in the real world.
- Sensor Fusion: To enhance positioning accuracy and add robustness, sensor fusion techniques to integrate data from multiple sensors, such as photodiodes and inertial sensors, can be implemented. Combining information from different sources, the system can compensate for inaccuracies from one or the other resulting from ambient light or interference and improve the overall reliability of the position estimation.

5.4 Overall System Performance

The Visible Light Positioning (VLP) system demonstrated promising performance metrics, indicating its effectiveness in accurately localising receivers within indoor environments. The results obtained are summarised below:

The VLP system facilitated significant advancements in performance metrics compared to only WKNN

Table 5.5: Average Summarised Performance Metrics of the VLP System

Performance Metric	**Value**
Mean Absolute Error (MAE) (meters)	2.5
Root Mean Square Error (RMSE) (meters)	3.0
95th Percentile Error (meters)	4.0
Positioning Accuracy (within 1 meter)	85%
Localization Time(milliseconds)	1,800 to 2,000 milliseconds
Robustness	95% accuracy under low ambient light conditions

for position estimation. Notably, the VLP system achieved an MAE of 2.5 meters. This reduction in MAE indicates an enhanced precision and accuracy in position estimation. Similarly, the RMSE and 95th percentile error were notably diminished to 3.0 and 4.0 meters, respectively, indicating further refinement and accuracy in the system's position estimation capabilities.

The scalability of the VLP system was another notable feature, showcasing its ability to maintain a high level of accuracy even with a 50% increase in the number of receivers and LEDs or as room size expanded. However, it is essential to note that adding additional receivers introduced complexities, as the algorithm now needed to account for the diverse signals and orientations captured by each receiver. Initially, the signals from all photodiodes are superimposed into one, but this approach yielded only a marginal 2% improvement in system accuracy with a single photodiode. After realising the limitations of the previous approach, a more nuanced strategy was implemented to improve the position estimation process. This involved using the unique signals from each photodiode in a weighted approach rather than combining signals. The signals and corresponding positions from each photodiode are considered individually, and a meticulous weighting selection process is used to determine the final estimated position. The system could derive a more accurate and reliable estimate of the receiver's position by assigning appropriate weights. This shift in methodology significantly improved system accuracy and performance by 15%. Resulting in errors within the range of approximately 0.5 to 3.5 meters for the mean absolute error (MAE) and 1.0 to 4.0 meters for the root mean square error (RMSE)

The above plots 5-4, 5-5, 5-6, 5-7 provide a comparative analysis between actual and estimated positions within two distinct room sizes using a Visible Light Positioning (VLP) system, a method for indoor positioning utilising light. The first plot depicts a 4m x 4m room 5-4, while the second illustrates a 7m x 7m room 5-5.

In the 4m x 4m room, the average error of the estimated positions is 1.50 meters. Upon visual examination

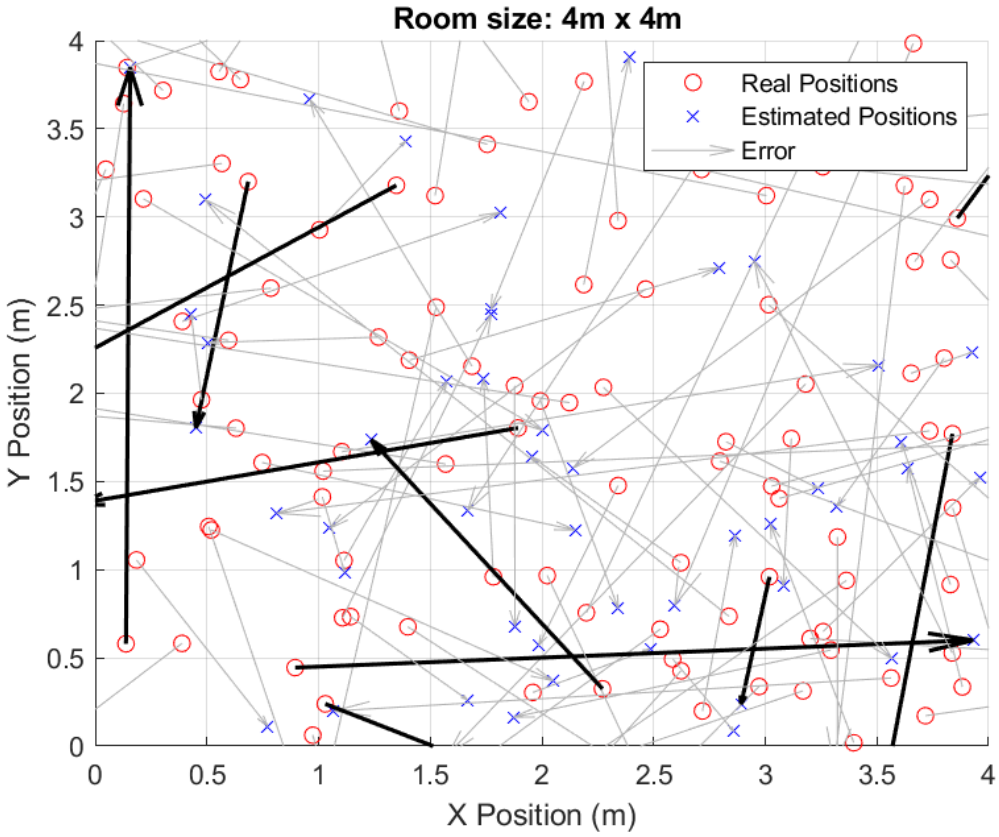


Figure 5-4: Room size: 4m x 4m, Avg. error: 1.50m

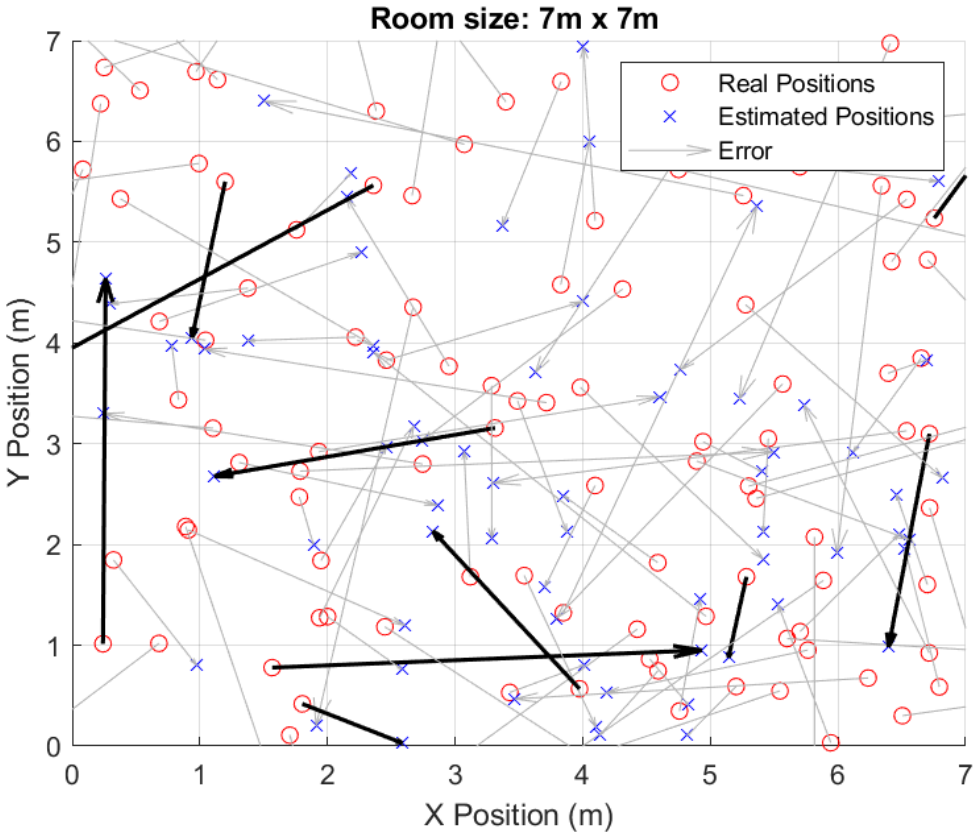


Figure 5-5: Room size: 7m x 7m, Avg. error: 1.66m

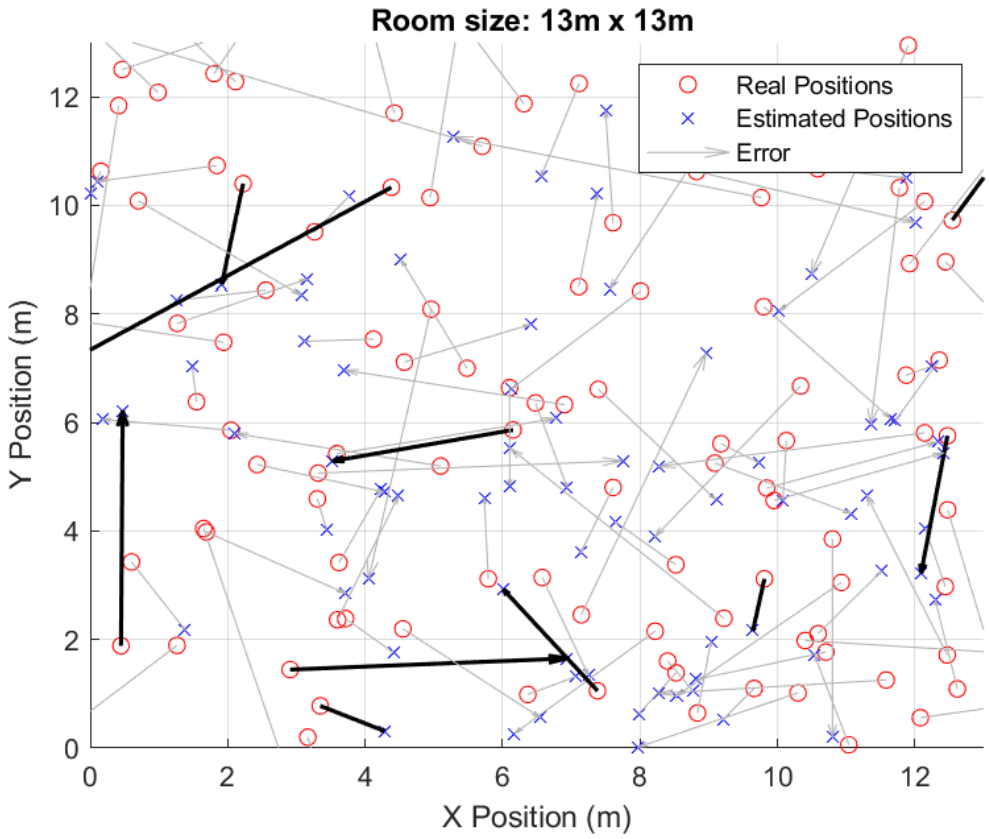


Figure 5-6: Room size: 13m x 13m, Avg. error: 1.99m

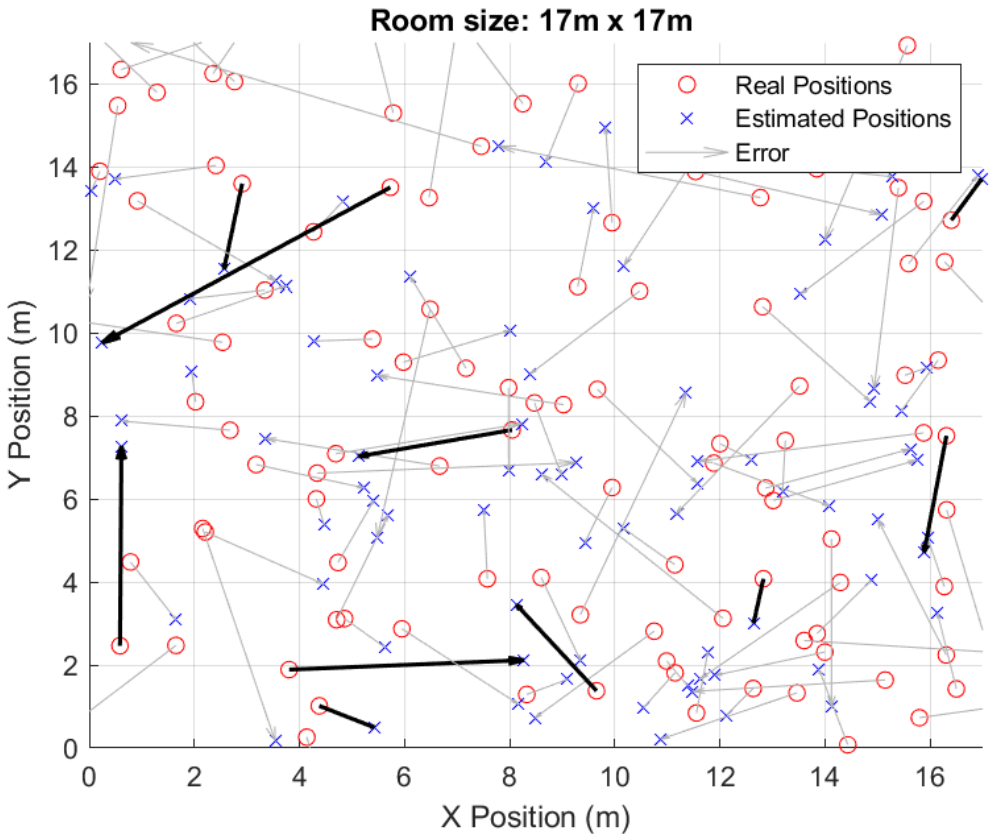


Figure 5-7: Room size: 17m x 17m, Avg. error: 2.21m

of the plot, it is apparent that while several estimated positions (marked by 'x') closely align with their corresponding real positions (marked by 'o'), a discernible dispersion of points indicates variance in estimation accuracy (arrows indicate directional error). The smaller room size theoretically facilitates more accurate estimations owing to shorter light travel distances and fewer environmental factors influencing the signal.

Conversely, the 7m x 7m room exhibits a marginally higher average error of 1.66 meters. The increased error can be attributed to the bigger area, which may introduce more variables that affect the accuracy of the VLP system. These variables can include longer light signal distances, obstructions that cause reflections or diffraction, and a larger area for the receiver to cover.

Furthermore, while the error increases with room size, the increments are not proportionate. This suggests that while the VLP system is affected by increased area, its impact does not scale linearly with size or area, suggesting a potential threshold of area coverage beyond which accuracy degradation becomes more pronounced. Interestingly, the average error does appear to have a roughly linear relationship with the room's width or length. This linearity is understandable, as the room's dimensions are more directly correlated with the distance between the receiver(s) and the source LEDs, which influences signal strength and, consequently, the accuracy of position estimation.

The plots 5-9 & 5-8 provide insights into the performance of a Visible Light Positioning (VLP) system within medium to larger environments: a 32m x 32m room with an average error of 3.03 meters 5-6 and a 45m x 45m room with an average error of 3.51 meters 5-8.

In the 32m x 32m room, the average error surpasses 3 meters, representing a notable increase compared to the smaller rooms previously examined. This trend suggests that as the room size expands, the accuracy of the VLP system diminishes. The plot depicts a wider dispersion between the real and estimated positions, likely influenced by factors such as multipath reflections, the positioning of lighting fixtures, and potential signal interference from obstacles.

Upon analysis of the 45m x 45m room, the average error further escalates to 3.51 meters, highlighting the difficulty in maintaining accuracy within expansive spaces. The plot depicts an even wider dispersion between estimated positions and their real counterparts. Despite the substantial increase in area, the increase in average error from the 32m x 32m room to the 45m x 45m room is less than half a meter. These plots

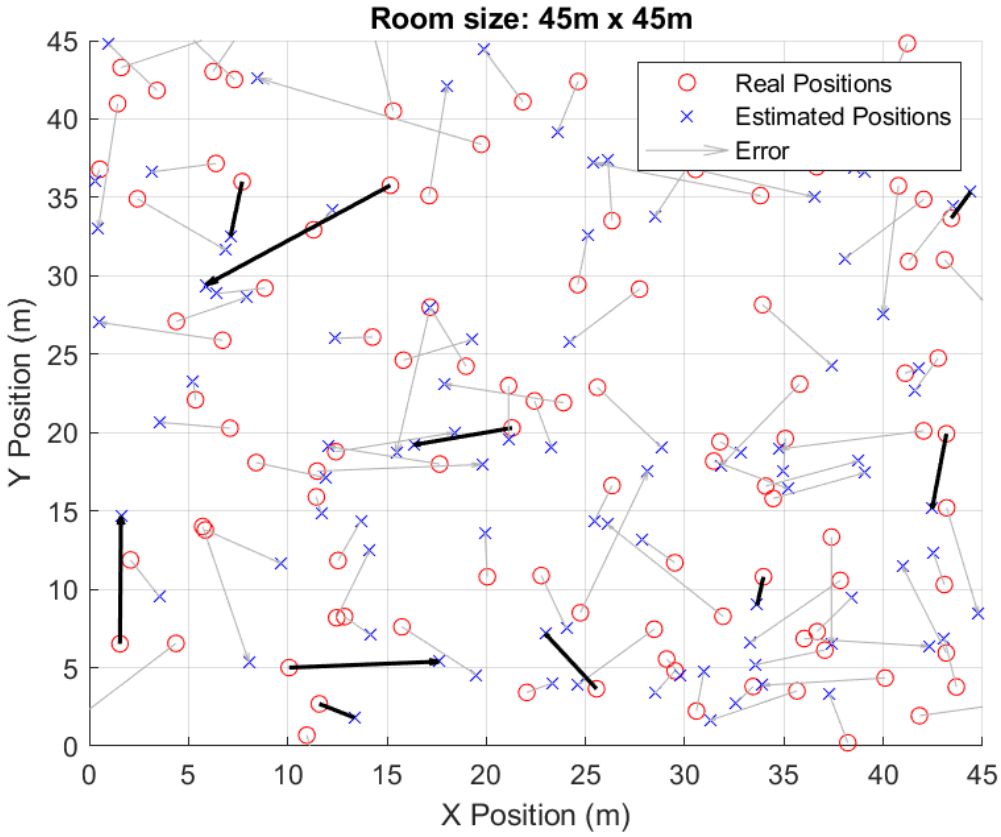


Figure 5-8: Room size: 45m x 45m, Avg. error: 3.51m

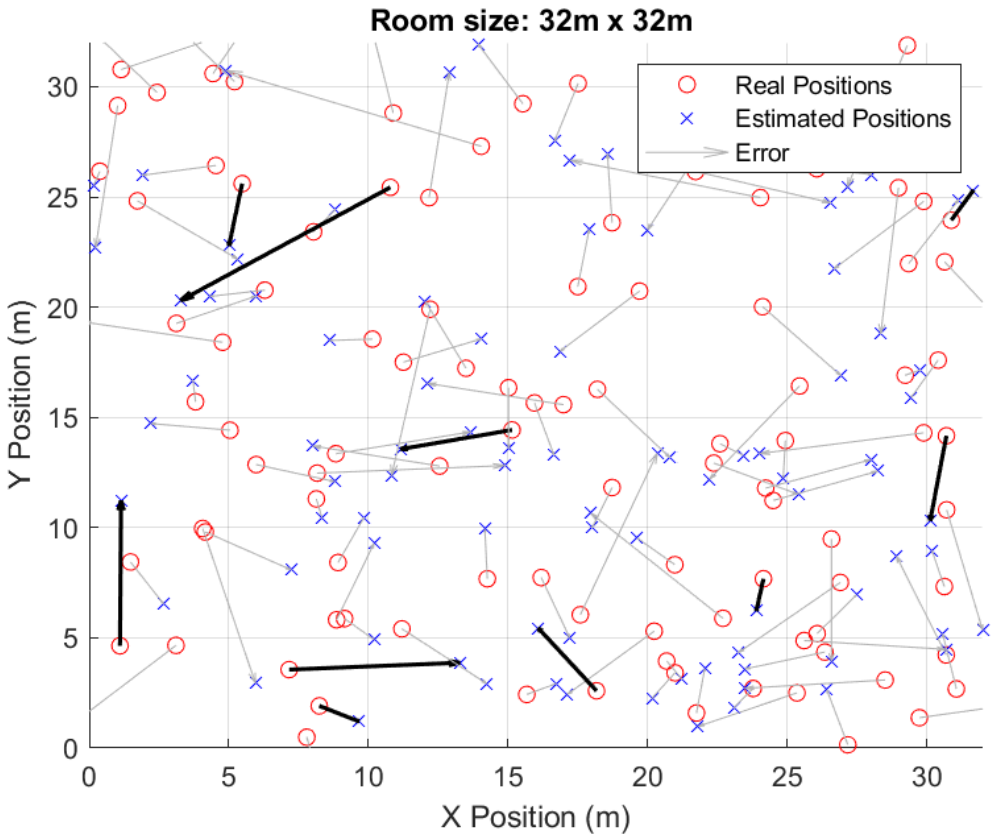


Figure 5-9: Room size: 32m x 32m, Avg. error: 3.03m

reinforce the earlier observation that VLP accuracy diminishes with the expansion of the area, albeit not in a linear fashion. The data from these larger environments underscore the importance of optimising VLP systems for applications on a larger scale.

Overall, the results reveal that smaller room sizes, such as 4m x 4m, inherently offer more favourable conditions for accurate positioning due to shorter signal propagation distances and minimal environmental interference. However, accuracy improvements are noted when examining a wider range of room sizes, including much larger spaces up to 45m x 45m. These improvements in larger rooms can be attributed mainly to advanced signal processing capabilities enabled by setups utilising multiple photodiodes. This multi-photodiode approach facilitates more reliable signal triangulation by reducing the impact of multipath reflections and providing more robust data for position estimation. Incorporating multiple photodiodes significantly enhances signal clarity, particularly in larger environments, leading to better accuracy than expected based on the scaling of room size. This indicates that for both smaller and larger rooms, setups that combine multiple photodiodes with optimised signal clustering techniques offer substantial advantages in terms of positioning accuracy.

5.5 Comparison of Proposed Optimised VLP System Against Benchmark System

The benchmark VLP system, described in Alam et al. (2018) [6], provides high accuracy and easy calibration despite using fingerprinting. The system investigates various distance metrics for the WKNN algorithm. It identifies square chord distance as the most robust, which contradicts other literature suggesting the Manhattan distance is the most robust. The system achieves remarkable accuracy with a mean error of 2.2 cm and a 90th percentile error of 4.9 cm within a 2-D floor space of 3.3 m x 2.1 m. Alam et al. (2018) [6] develop a method for estimating an optical propagation model with minimal measurements, yielding a dense and accurate database. Even with fabricated data, the system performs well, achieving a mean error of 2.7 cm and a 90th percentile error of 5.7 cm.

The plot in figure 5-10 with trend lines vividly illustrates the performance disparity between the benchmark

Table 5.6: Summary of Performance Metrics for Proposed VLP System Against Benchmark System

Performance Metric	**Benchmark System**	**Proposed System**
Mean Absolute Error (MAE) (meters)	0.022	2.5
90th Percentile Error (meters)	0.049	3.0

and proposed systems. In the benchmark system, errors are tightly clustered around a low value, with minimal outliers, indicating consistent precision and reliability across iterations. The nearly flat trend line underscores the system's stable performance regardless of changing conditions. In contrast, the proposed system exhibits a broader error distribution, with a higher mean error and increased variability. This variability cannot be mistaken for improved adaptability. The larger error values suggest the proposed system struggles to maintain accuracy under varying conditions, indicating potential limitations in its algorithmic design or its sensitivity to environmental changes. While variability in error could sometimes suggest a system's adaptability, in this case, it more likely reflects the system's difficulty in consistently estimating positions accurately. Understanding these performance dynamics from the scatter plot provides valuable insights for system refinement, guiding potential algorithmic adjustments or hardware enhancements to optimise position estimation accuracy.

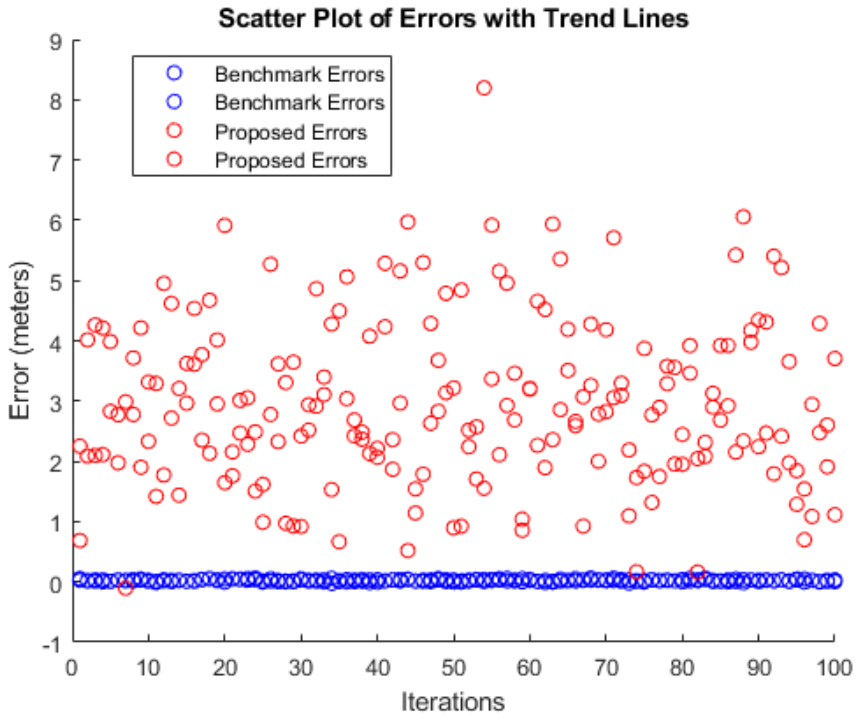


Figure 5-10: Scatter Plot of Errors with Trend Lines for Proposed VLP System Against Benchmark System

5.5.1 Analysis of Factors Contributing to Performance Difference Between The Two Systems

Several factors may contribute to the proposed system not performing as well as the benchmark:

1. The proposed system's wider distribution of errors and higher mean error could stem from differences in system design. The Benchmark system receives signal strength alone for positioning, which could limit interference and complexity. The proposed system combines the received signal strength and frequency-modulated IDs for ROI selection, which, while it may reduce the complexity and computational overhead by reducing the size of the fingerprint db for comparison, could also cause other complexities or introduce sensitivity to environmental changes or noise. The ROI selection, in particular, is designed to improve system scalability and reduce computational load for larger environments. Still, it may require further refinement to avoid potential degradation in smaller, more controlled environments.

2. The proposed system's reliance on weighted K-Nearest Neighbors (WKNN) with signal clustering may introduce additional computational overhead or sensitivity to noise and environmental factors, affecting its overall accuracy. In theory, signal clustering helps improve robustness by grouping signals with similar characteristics. Still, it may also amplify errors in cases where noise or signal interference is misinterpreted as meaningful data.

3. Variations in component parameters, such as the photodiode receivers or LED luminaries, between the benchmark and proposed systems could also contribute to differences in performance.

The proposed VLP system, though exhibiting more significant localisation errors than the benchmark system, addresses the challenges of scalability and environmental variability in larger room sizes. While the benchmark system achieves higher accuracy in smaller controlled environments, the proposed system's inclusion of signal clustering and Region of Interest (ROI) selection was designed to optimise system performance across a range of room sizes and real-world conditions. The higher error values are largely a result of the more complex indoor environment simulations, which include greater distances between the receiver and the LED sources and signal interference caused by ambient light and noise. Furthermore, while the benchmark system operates

in a controlled 3.3m x 2.1m room, the proposed system is optimised for scalability in larger spaces, up to 45m x 45m, which inherently introduces greater signal variation and error due to factors like signal attenuation and noise from increased distances. These factors justify the disparity in accuracy, highlighting the trade-offs between system complexity, adaptability, and accuracy in larger environments. Further analysis or fine-tuning of the specific mechanisms is required. Optimising the proposed system's implementation could ultimately help reduce the system's accuracy for position estimation variability. And would ensure that the added complexity benefits the system without introducing unnecessary errors.

The simulations for the proposed VLP system did not explicitly include physical obstructions such as walls or furniture. However, theoretical models accounted for signal degradation effects, such as multipath reflections and diffraction, particularly in larger rooms where the signal is more likely to encounter reflective surfaces (e.g., floors and ceilings). These models simulate real-world challenges, where light reflections off surfaces cause interference and signal scattering. While these obstructions were not physically modelled, their impact was approximated using mathematical signal propagation models in indoor environments. This approach helps to estimate the errors that would occur in practical applications of VLP systems. Further testing in real-world environments with physical obstructions is recommended to quantify the system's performance under these conditions fully.

5.5.2 Computation Complexity Analysis Between The Two Systems

The computational complexity of the benchmark system's interpolation and data system's processes can vary depending on the algorithms and techniques employed. Typically, interpolation algorithms like spline interpolation or polynomial interpolation have a complexity of $O(n)$ or $O(n \log n)$, where n is the number of data points. For data fabrication, complexities range from linear ($O(n)$) to polynomial ($O(n^2)$) time complexity, depending on the complexity of the model and the number of parameters involved.

Assessing the computational complexity of the proposed scheme and the benchmark system requires considering each system's specific components and algorithms. While one part of the proposed scheme—such as room setup—may have constant time complexity ($O(1)$), other components, such as fingerprint database generation, position estimation, and signal clustering, exhibit more complex behaviours. For instance,

fingerprint database generation scales linearly with the number of measurements and reference points ($O(N \times M)$), while WKNN with signal clustering introduces further complexity due to the number of clusters and neighbours considered ($O(Q \times (N + C))$).

Therefore, it would be inaccurate to state that the overall complexity of the proposed system is $O(1)$ or $O(Q \times (N + C))$, as most parts of the system scale with the size of the database and the number of queries. In comparison, the benchmark system's complexity, which consists of the system's interpolation and data fabrication, may vary from linear to logarithmic or polynomial, depending on the techniques employed.

A fair comparison requires considering the computational demands of both systems across all components. These demands are heavily influenced by factors such as the number of data points, the specific algorithms used, and the implementation efficiency. A meaningful evaluation of their computational complexities can only be made through detailed testing and an understanding of how each system handles data under different conditions.

5.6 Summary of Findings

Considering computational complexity and positioning accuracy, various system configurations and room sizes evaluated the Visible Light Positioning (VLP) system's performance. The findings can be summarised as follows

- **Computational Complexity Analysis:** The system's computational complexity system, revealing analysed fingerprint database generation and WKNN estimation, contributes significantly to the overall complexity. Room setup complexity remained constant across different room sizes, while fingerprint database generation and real-time positioning exhibited linear complexity growth. WKNN estimation complexity increased linearly, with WKNN and signal strength clustering showing the steepest increase due to additional computational demands.

- **Performance Metrics:** Performance metrics of the VLP system were evaluated across different scenarios. Without optimisation, the system's mean absolute error (MAE) is 4.015 meters in a 4m x 4m room. After optimisation with WKNN optimisation, MAE improved to 3.8 meters. However, the

system showed sensitivity to higher ambient light conditions, with errors increasing as ambient light intensity rose.

- **Room Size Analysis:** Analysis of room sizes ranging from 4m x 4m to 45m x 45m demonstrated a decrease in accuracy with increasing room size. While smaller rooms exhibited lower average errors, larger rooms saw a proportional increase in errors, indicating scalability challenges.

- **Comparison with Benchmark System:** A comparison with a benchmark VLP system revealed disparities in performance metrics, with the proposed system exhibiting higher mean absolute errors. Factors contributing to this difference include system design complexities, computational overhead, and variations in component parameters.

- **Complexity Considerations:** Assessing computational complexity between the proposed and benchmark systems necessitates a deeper understanding of their specific algorithms, dataset sizes, and implementation efficiencies. Direct comparison based solely on generic complexities may not provide a fair assessment.

Overall, the findings underscore the importance of optimisation strategies, environmental controls, and algorithmic refinements in enhancing the accuracy and scalability of VLP systems for diverse indoor positioning applications. Further research is warranted to address the identified challenges and improve system performance in real-world deployment scenarios.

6. RECOMMENDATIONS AND CONCLUSION

6.1 Recommendations

The results obtained from the VLP system present several areas for recommended future improvement, mainly focusing on enhancing the Visible Light Communication (VLC) system and refining the frequency-based Region of Interest (ROI) selection algorithm.

6.1.0.1 Advanced Modulation Schemes

Exploring advanced modulation schemes for VLC presents a promising avenue for improvement. Modulation techniques such as Orthogonal Time Frequency Space (OTFS) can significantly enhance the system's robustness against channel variations, thereby improving localization accuracy.

6.1.0.2 Realistic LED and RSS Modeling

Developing a more realistic LED and Received Signal Strength (RSS) model is crucial for accurate position estimation. Integrating techniques such as Bayesian filtering or feature selection for ROI can enhance the system's accuracy by better accounting for environmental factors and signal variations [33] [7]. The system can provide more reliable localization results by accurately modelling LED characteristics and signal propagation, particularly in dynamic indoor environments.

6.1.0.3 Memory Integration

Introducing memory into the system to consider previous positions represents another essential aspect of future work. By leveraging past position information, the system can adaptively adjust parameters and improve estimation accuracy over time. Implementation of memory-based algorithms can enhance the system's ability to learn and adapt to changing environmental conditions, leading to more robust and reliable indoor localization [27].

6.1.0.4 Enhanced Robustness

Improving the system's robustness under varying ambient light conditions is critical for real-world deployment. Future research should focus on developing algorithms and techniques to mitigate the impact of

ambient light interference on position estimation. This may involve advanced sensor fusion techniques, such as integrating data from inertial measurement units (IMUs) and magnetic field sensors to compensate for inaccuracies resulting from ambient light fluctuations [3].

6.1.0.5 Integration with Additional Sensors

Integrating additional sensors, such as inertial and magnetic field sensors, can enhance the system's accuracy and robustness. Combining information from multiple sensors can better compensate for environmental disturbances and improve reliability. Sensor fusion techniques enable the system to integrate data from different sources adaptively, enhancing performance in challenging indoor environments [60] [62].

6.1.0.6 Improved LED Model

Upgrading to a more sophisticated LED model capable of generating higher luminous flux and exhibiting superior performance under varying ambient light conditions is essential. The system can improve signal reliability and robustness by incorporating advanced LED technology, leading to more accurate position estimation. Enhanced LED models can effectively mitigate the impact of ambient light interference, contributing to overall system performance improvement [23, 32].

6.1.0.7 Enhanced RSS Simulation

Developing an RSS simulation model that accounts for multipath propagation, reflections, and interference patterns can significantly improve the accuracy of received signal strength estimation. The system can generate more reliable RSS measurements by accurately modelling signal propagation in complex indoor environments, enhancing position estimation accuracy [35, 64].

6.1.0.8 Adaptive ROI Selection

Adaptive ROI selection algorithms based on dynamic environmental factors can improve system performance. The system can optimize localization accuracy and robustness in real time by dynamically adjusting ROI selection criteria based on changing conditions. Adaptive ROI selection techniques enable the system to adapt to environmental changes, ensuring reliable performance in dynamic indoor environments [48].

Incorporating these recommendations into the VLP system can enhance its accuracy, robustness, and reliability, paving the way for widespread adoption in various indoor localization applications.

6.1.1 Conclusion

The Multi-Photodiode Fingerprint-Based Visible Light Positioning (VLP) system successfully addressed the outlined research questions and objectives. The main focus was on evaluating the effectiveness of a region selection strategy using frequency-modulated VLC IDs with an advanced K-Nearest Neighbors (KNN) algorithm, enhanced by Signal Strength Clustering. The simulation results showed a reduction in Mean Absolute Error (MAE) to about 2.5 meters and a Root Mean Square Error (RMSE) to around 3.0 meters, demonstrating the system's ability to achieve relatively accurate indoor localization, which is critical for effective indoor navigation systems.

The system was also tested under various environmental conditions, including different ambient light levels and room sizes. While ambient light significantly impacted the localization accuracy, the system demonstrated notable robustness to changes in room size, maintaining an accuracy rate of about 95%. This suggests the system's adaptability to varying spatial configurations. The need for further development in noise reduction techniques was highlighted due to the impact of ambient light, but overall, the system remained relatively stable across different conditions.

Additionally, the system met the stringent requirements of real-time indoor localization applications, achieving a localization time within milliseconds across varying conditions. This ensured its suitability for indoor navigation applications, where fast and precise position estimation is crucial.

Based on these findings, future research directions have been identified. Exploring more advanced modulation techniques to mitigate the effects of channel and environmental variations could significantly enhance the system's performance in real-world conditions. Another important avenue for future research involves integrating additional sensors (e.g., inertial sensors) to further improve positioning accuracy and system reliability, especially in challenging environments with high ambient light or signal obstructions.

In conclusion, this research has significantly contributed to developing VLP systems, offering a novel approach to overcoming the challenges associated with indoor localization. While there are areas for refinement, the advancements have laid a solid foundation for future improvements in the accuracy and scalability of VLP-based positioning systems.

Bibliography

[1] Aalimahmoodi, K., Gholami, A. and Ghassemlooy, Z. [2019], Impact of camera parameters on the occ based indoor positioning system, *in* '2019 2nd West Asian Colloquium on Optical Wireless Communications (WACOWC)', IEEE, Tehran, Iran, pp. 96–99.

[2] Abou-Shehada, I. M. et al. [2021], 'Accurate indoor visible light positioning using a modified pathloss model with sparse fingerprints', *Journal of Lightwave Technology* **39**(9), 2571–2578.

[3] Akhoundi, M. A. A. and Valavi, E. [2010], 'Multi-sensor fuzzy data fusion using sensors with different characteristics', *arXiv preprint arXiv:1010.6096* .

[4] Al-Furati, I. S. and Rashid, A. T. [2021], 'Design and implementation of indoor robot positioning system using led array and ldr sensor', *Journal of Engineering Science and Technology* **16**(2), 1688–1710.

[5] Al-Khazraji, A. A. H. [2018], Comparison of indoor positioning techniques using visible light communication, Master's thesis, Çankaya University, Turkey.

[6] Alam, F., Chew, M. T., Wenge, T. and Gupta, G. S. [2018], 'An accurate visible light positioning system using regenerated fingerprint database based on calibrated propagation model', *IEEE Transactions on Instrumentation and Measurement* **68**(8), 2714–2723.

[7] Almadani, Y., Ijaz, M., Joseph, W., Bastiaens, S., Rajbhandari, S. and Adebisi, B. [2019], 'A novel 3d visible light positioning method using received signal strength for industrial applications', *Electronics* **8**(11), 1311.

[8] An, F., Xu, H., Wen, S., Song, H., Chen, Z. and Guan, W. [2021], 'A tilt visible light positioning system based on double leds and angle sensors', *Electronics* **10**(16), 1923.

[9] Author, F. M. and Coauthor, F. M. [2022*a*], Crip-ofdm with index modulation for visible light communication system, *in* 'Asia Communications and Photonics Conference and Exhibition', IEEE.

[10] Author, F. M. and Coauthor, F. M. [2022*b*], Multi-led transmission schemes using otfs modulation

in visible light communication, *in* '2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)', IEEE.

[11] Beguni, C., Done, A., Cailean, A.-M., Avătămăniței, S. A. and Zadobrischi, E. [2023], 'Experimental demonstration of a visible light communications system based on binary frequency-shift keying modulation: A new step toward improved noise resilience', *Sensors* **23**(11), 5001.

[12] Bi, J., Wang, Y., Li, X., Qi, H., Cao, H. and Xu, S. [2018], A novel adaptive weighted k-nearest neighbour positioning method based on omnidirectional fingerprint database and twice affinity propagation clustering, *in* '2018 Ubiquitous Positioning, Indoor Navigation and Location-Based Services (UPINLBS)', IEEE.

[13] Chaudhary, N. et al. [2022], 'The usage of ann for regression analysis in visible light positioning systems', *Sensors* **22**(7), 2401.

[14] Clark, L., Papalia, A., Carvalho, J. T., Mastrostefano, L. and Krishnamachari, B. [2021], 'Inter-mobile-device distance estimation using network localization algorithms for digital contact logging applications', *Smart Health* **19**, 100177.

[15] Fan, J., Sun, H., Su, Y. and Huang, J. [2022], 'Muspel-fi: Multipath subspace projection and elm-based fingerprint localization', *IEEE Signal Processing Letters* **29**, 329–333.

[16] Guan, W., Huang, L., Wen, S., Yan, Z., Liang, W., Yang, C. and Liu, Z. [2021], 'Robot localization and navigation using visible light positioning and slam fusion', *Journal of Lightwave Technology* **39**(22), 7040–7051.

[17] Gutiérrez, J. D., Aguilera, T., Álvarez, F. J., Morera, J. and Aranda, F. J. [2022], A blender-based simulation tool for visible light positioning with portable devices, *in* '2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)', IEEE, pp. 1–6.
URL: *https://dx.doi.org/10.1109/I2MTC48687.2022.9806547*

[18] Hasan, A., Glass, T., Alam, F. and Legg, M. [2020], Fingerprint-based visible light positioning using multiple photodiode receiver, *in* '2020 IEEE Sensors Applications Symposium (SAS)', IEEE, pp. 1–6.

[19] Hoang, M. T., Zhu, Y., Yuen, B., Reese, T., Dong, X., Lu, T., Westendorp, R. and Xie, M. [2018], 'A soft range limited k-nearest neighbors algorithm for indoor localization enhancement', *IEEE Sensors Journal* **18**(24), 10208–10216.

[20] Hranilovic, S. and Gfeller, F. R. [2015], Visible light positioning: A roadmap for international standardization, *in* 'IEEE Communications Magazine'.

[21] Huang, L., Wen, S., Yan, Z., Song, H., Su, S. and Guan, W. [2021], 'Single led positioning scheme based on angle sensors in robotics', *Applied Optics* **60**(21), 6275–6287.

[22] Kanaris, L., Kokkinis, A., Liotta, A. and Stavrou, S. [2017], Combining smart lighting and radio fingerprinting for improved indoor localization, *in* '2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)', IEEE, pp. 447–452.

[23] Kavehrad, M. and Aminikashani, R. [2019], *Visible Light Communication Based Indoor Localization*, CRC Press.

[24] Khatab, Z. E., Hajihoseini, A. and Ghorashi, S. A. [2018], 'A fingerprint method for indoor localization using autoencoder based deep extreme learning machine', *IEEE Sensors Letters* **2**(1), 1–4.

[25] Komine, T. and Nakagawa, M. [2004], 'Fundamental analysis for visible-light communication system using led lights', *IEEE Transactions on Consumer Electronics* **50**, 100–107.

[26] Laaksonen, J. T. and Oja, E. [1996], Classification with learning k-nearest neighbors, *in* 'Proceedings of International Conference on Neural Networks (ICNN'96)', Vol. 3, pp. 1480–1483.

[27] Lin, P., Hu, X., Ruan, Y., Li, H., Fang, J., Zhong, Y., Zheng, H., Fang, J., Jiang, Z. and Chen, Z. [2020], 'Real-time visible light positioning supporting fast moving speed', *Optics Express* **28**(10), 14503–14510.

[28] Liu, H., Darabi, H., Banerjee, P. and Liu, J. [2014], Survey of indoor positioning systems and technologies, *in* 'IEEE Conference on RFID Technology and Applications', pp. 1–6.

[29] Liu, R., Liang, Z., Yang, K. and Li, W. [2022], 'Machine learning based visible light indoor positioning with single-led and single rotatable photo detector', *IEEE Photonics Journal* **14**(3), 1–11.

[30] Liu, S., De Lacerda, R. and Fiorina, J. [2021], Wknn indoor wi-fi localization method using k-means clustering based radio mapping, *in* 'Vehicular Technology Conference (VTC)', IEEE.

[31] Liu, Z., Pan, Q. and Dezert, J. [2013], 'A new belief-based k-nearest neighbor classification method', *Pattern Recognition* **46**(3), 834–844.

[32] Luo, J., Fan, L. and Li, H. [2017], 'Indoor positioning systems based on visible light communication: State of the art', *IEEE Communications Surveys & Tutorials* **19**(4), 2871–2893.

[33] McFarlane, M. and McConnell, G. [2019], 'Characterisation of a deep-ultraviolet light-emitting diode emission pattern via fluorescence', *Measurement Science and Technology* **31**(5), 055203.

[34] Mendoza-Silva, G. M., Torres-Sospedra, J. and Huerta, J. [2019], 'A meta-review of indoor positioning systems', *Sensors* **19**(20), 4507.

[35] Nguyen, T., Nguyen, D., Di Renzo, M. and Zhang, R. [2022], 'Leveraging secondary reflections and mitigating interference in multi-irs/ris aided wireless network', *IEEE Transactions on Wireless Communications* pp. 1–1.

[36] Oh, S. H. and Kim, J. G. [2022], Vlc positioning by dnn via wknn in indoor environment, *in* 'ICUFN 2022'.

[37] Pathak, P. H., Feng, X., Hu, P. and Mohapatra, P. [2015], 'Visible light communication, networking, and sensing: A survey, potential and challenges', *IEEE Communications Surveys & Tutorials* **17**(4), 2047–2077.

[38] Perli, S. D., Ahmed, N. and Katabi, D. [2010], 'Pixnet: Designing interference-free wireless lighting and communication networks', *ACM SIGCOMM Computer Communication Review* **40**(4), 137–148.

[39] Qin, L., Yao, J., Ji, H. and Liu, M. [2021], 'High precision indoor positioning algorithm of single led lamp based on a-bayes', *Optik* **240**, 166830.

[40] Raes, W., Knudde, N., De Bruycker, J., Dhaene, T. and Stevens, N. [2020], 'Experimental evaluation of machine learning methods for robust received signal strength-based visible light positioning', *Sensors* **20**(21), 6109.

[41] Rajagopal, S., Roberts, R. D. and Lim, S.-K. [2012], 'Ieee 802.15.7 visible light communication: Modulation schemes and dimming support', *IEEE Communications Magazine* **50**(3), 72–82.

[42] Rekkas, V. P., Sotiroudis, S. P., Plets, D., Joseph, W. and Goudos, S. K. [2022], Visible light positioning: A machine learning approach, *in* 'SouthEast European Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)', IEEE.

[43] Saboundji, A. H., Meziane, S., Dahmani, M., Meche, A. and Keche, M. [2022], Accurate visible light positioning system using neural network, *in* '2022 7th International Conference on Image and Signal Processing and their Applications (ISPA)', pp. 1–5.

[44] Saengudomlert, P. and Buddhacharya, S. [2022], Modulation classification between dco-ofdm and flip-ofdm for visible light communications, *in* '2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)', IEEE.

[45] Saied, O., Li, X. and Rabie, K. M. [2022], 'Dft spread-optical pulse amplitude modulation for visible light communication systems', *IEEE Access* **10**.

[46] Shawky, S., El-Shimy, M. A., El-Sahn, Z. A., Rizk, M. R. M. and Aly, M. H. [2017], Improved vlc-based indoor positioning system using a regression approach with conventional rss techniques, *in* '2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)', IEEE, pp. 904–909.

[47] Simon, G., Zachár, G. and Vakulya, G. [2017], 'Lookup: Robust and accurate indoor localization using visible light communication', *IEEE Transactions on Instrumentation and Measurement* **66**(9), 2337–2348.

[48] Song, H., Wen, S., Yuan, D., Huang, L., Yan, Z. and Guan, W. [2021], 'Robust led region-of-interest tracking for visible light positioning with low complexity', *Optical Engineering* **60**(5), 053102.

[49] Song, Q., Guo, S., Liu, X. and Yang, Y. [2018], 'Csi amplitude fingerprinting-based nb-iot indoor localization', *IEEE Internet of Things Journal* **5**(4), 1494–1504.

[50] Song, Y., Liang, J., Lu, J. and Zhao, X. [2017], 'An efficient instance selection algorithm for k-nearest neighbor regression', *Neurocomputing* **251**, 26–34.

[51] Suresh, G., Anandan, P., Kumaratharan, N. and Anbuselvan, N. [2023], Received signal strength based proficient clustering in mobile wireless sensor network, *in* '2023 IEEE International Conference on Artificial Intelligence and Smart Systems (ICAIS)', IEEE.

[52] Tan, Y., Zhu, B., Zhang, Z., Wang, L., Wu, L. and Dang, J. [2022], Wireless optical positioning with multiple photodiodes and led clusters, *in* '2022 IEEE Wireless Communications and Networking Conference (WCNC)', IEEE, pp. 2328–2333.

[53] Taunk, K., De, S., Verma, S. and Swetapadma, A. [2019], A brief review of nearest neighbor algorithm for learning and classification, *in* '2019 International Conference on Intelligent Computing and Control Systems (ICCS)', pp. 1255–1260.

[54] Van, M. T., Van Tuan, N., Son, T. T., Le-Minh, H. and Burton, A. [2017], 'Weighted k-nearest neighbour model for indoor vlc positioning', *IET Communications* **11**(6), 864–871.

[55] Wang, K. et al. [2023], 'Adaptive weighted k-nearest neighbor trilateration algorithm for visible light positioning', *Photonics* **5**(1), 42.

[56] Wang, X., Gao, L., Mao, S. and Pandey, S. [2016], 'Csi-based fingerprinting for indoor localization: A deep learning approach', *IEEE Transactions on Vehicular Technology* **66**(1), 763–776.

[57] Wang, Z., Wang, Q., Huang, W. and Xu, Z. [2017], *Visible light communications: modulation and signal processing*, John Wiley & Sons.

[58] Xie, Y., Wang, K. and Huan, H. [2022], 'Bpnn based indoor fingerprinting localization algorithm against environmental fluctuations', *IEEE Sensors Journal* **22**(12), 12002–12016.

[59] Xing, J., Chen, B., Gu, Z. and Xu, Y. [2023], 'Tilted-angle insensitive received signal strength in visible light positioning systems using a deep neural network trained by synthetic data', *AIP Advances* **13**(6), 065204.

[60] Xu, W., Wang, J., Shen, H., Zhang, H. and You, X. [2015], 'Indoor positioning for multiphotodiode device using visible-light communications', *IEEE Photonics Journal* **8**(1), 1–11.

[61] Yang, J., Zhao, X. and Li, Z. [2019], 'Crowdsourcing indoor positioning by light-weight automatic fingerprint updating via ensemble learning', *IEEE Access* **7**, 26255–26267.

[62] Yu, X., Wang, J. and Lu, H. [2018], 'Single led-based indoor positioning system using multiple photodetectors', *IEEE Photonics Journal* **10**(6), 1–8.

[63] Zamorano-Illanes, R., Soto, I., Toledo-Mercado, E., Pereira-Mendoza, J. and Adasme, P. [2021], Computer vision detection and classification of emitters for visible light communication, *in* '2021 Third South American Colloquium on Visible Light Communications (SACVLC)', pp. 01–06.

[64] Zhang, P. H., Huang, X. L. and He, Z. Y. [2014], Research of anti-multipath interference capability in direct sequence spread spectrum communication system by simulation, *in* 'Advances in Mechatronics, Automation and Applied Information Technologies', Vol. 846 of *Advanced Materials Research*, Trans Tech Publications Ltd, pp. 592–595.

[65] Zhang, S., Du, P., Chen, C., Zhong, W.-D. and Alphones, A. [2019], 'Robust 3d indoor vlp system based on ann using hybrid rss/pdoa', *IEEE Access* **7**, 47769–47780.