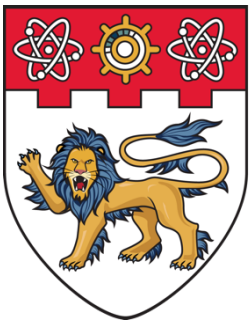


NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

SMARTPHONE AIDED CMOS SENSOR BASED INDOOR VISIBLE LIGHT POSITIONING TECHNIQUES

SRIVATHSAN CHAKARAVARTHI NARASIMMAN

School of Electrical & Electronic Engineering

2025

SMARTPHONE AIDED CMOS SENSOR BASED INDOOR VISIBLE LIGHT POSITIONING TECHNIQUES

SRIVATHSAN CHAKARAVARTHI NARASIMMAN

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy

2025

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research, is free of plagiarised materials, and has not been submitted for a higher degree to any other University or Institution.

16 Aug. 2024
.....
Date

C.N. Sri
.....
SRIVATHSAN CHAKARAVARTHI NARASIMMAN

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it is free of plagiarism and of sufficient grammatical clarity to be examined. To the best of my knowledge, the research and writing are those of the candidate except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

16 Aug. 2024

.....

Date

.....

Assoc. Prof. Arokiaswami Alphones

Authorship Attribution Statement

* This thesis contains material from 4 papers published or accepted in the following peer-reviewed journals/conferences in which I am listed as the first author.

Chapter 3 includes work published as S.C. Narasimman, and A. Alphones. "Indoor Visible Light Positioning for A Single Partially Visible LED" IEEE Sensors Letters, 1-4,,vol 8, issue 5, (2024). DOI: 10.1109/LSENS.2024.3385543.

The contributions of the co-authors are as follows:

- A/Prof Alphones provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript draft, collected and analysed experimental data.

Chapter 4 includes work presented as S.C. Narasimman, and A. Alphones. "Tree-based Single LED Indoor Visible Light Positioning Technique" TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON), 826-831 (2023).
DOI: 10.1109/TENCON58879.2023.10322350.

The contributions of the co-authors are as follows:

- A/Prof Alphones provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript draft, collected and analysed experimental and simulated data.

Chapter 5 includes work accepted as S.C. Narasimman, and A. Alphones. "Simulation and Experimental Validation of Optical Camera Communication" TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON),2024

The contributions of the co-authors are as follows:

- A/Prof Alphones provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript draft, collected and analysed experimental and simulated data.

Chapter 6 includes work published as S.C. Narasimman, and A. Alphones. "DumbLoc: Dumb Indoor Localization Framework using WiFi Fingerprinting" IEEE Sensors Journal, 14623-14630,vol 24, issue 9, (2024). DOI: 10.1109/JSEN.2024.3374415.

The contributions of the co-authors are as follows:

- A/Prof Alphones provided the initial project direction and edited the manuscript drafts.
- I prepared the manuscript draft, formulated a processing technique and analysed the results.

16 Aug. 2024

. . . . . .

Date

C.N. Sri

. . . . . .

SRIVATHSAN CHAKARAVARTHI NARASIMMAN

Acknowledgements

I wish to express my greatest gratitude to my supervisor Professor Arokiaswami Alphones, without whose patient guidance this would not have been possible.

I would like to thank the Surbana Jurong-Nanyang Technological University corporate laboratory for providing the scholarship and the infrastructure to conduct this research.

Thanks to my parents, my sister and family for the constant support.

Contents

Abstract

Indoor positioning systems (IPS) have been researched extensively both commercially and in academia owing to the wide array of applications they cater to such as indoor navigation, occupancy tracking, asset tracking, virtual or augmented reality, targeted advertisement or information delivery and point cloud registration. Several radio frequency (RF) applications have been employed in an attempt to solve this problem. However, due to interference and multi path effects, they have struggled to achieve high positioning accuracy and scalability. One technique gaining traction over the past decade is visible light communication (VLC), where the frequency of the carrier wave is in the visible light frequency range of the electromagnetic spectrum. The widespread adoption of light emitting diodes (LEDs) has led to a simultaneous reduction in price and improvement in quality which has pushed visible light positioning (VLP) to the fore since commercially available LEDs are capable of supporting high switching rates with improved reliability. While there are several extant positioning techniques, VLP has a unique set of advantages, such as higher accuracy and lack of interference, which makes it viable for further study.

VLP techniques employ photodetectors (PDs) or cameras, with the former playing a major role in trilateration and the latter often playing second fiddle. While it is true that with the standardization and widespread adoption of VLC, the PD may make its way on to more commercial devices this will take a decade at the earliest. The camera on a smartphone though is ready to use now and is here to stay. Of the popular camera sensors complementary metal oxide semiconductor (CMOS) performs better than charge coupled device (CCD) sensors in areas of power consumption, cost and frame rate making it the natural choice for VLP. The performance of the most popular indoor positioning techniques such as Wi-Fi based, magnetometer based or pedestrian dead reckoning approaches leverage sensors found on the modern smartphones. Hence, the smartphone should be at the center of an indoor positioning technique due to its ubiquity and ease of use. Therefore, CMOS sensor-based indoor VLP schemes with the smartphone at the center, leveraging the onboard sensors, are to be explored. CMOS sensor based VLP

techniques can be split into two parts. The first of which is localization of the user with respect to the transmitter and the second involves identifying the location of the transmitter. We explore different techniques to solve each of these parts of the VLP problem.

While several VLP techniques have seen success using CMOS sensors for a single LED, in most cases the field of view (FoV) of a front camera on a smartphone is much smaller than the rear camera and lights are placed sparsely in offices since their primary objective is illumination. Hence, during indoor navigation the front camera is bound to encounter far more partial images of the light than complete images. We proposed a technique to solve this problem by performing positioning on images where only two corners of a square light are in the FoV of the camera. While most offices have square or rectangular panel lights, we have chosen to use square lights owing to the added difficulty in positioning arising from all sides being equal in length. We detect the corners of the light from an image and order them based on inertial measurement unit (IMU) readings from smartphones to perform structure-based positioning.

Though we achieved high accuracy with a 3D geometry-based technique to identify the relative position of the receiver with respect to the transmitter, we tested the performance of machine learning models on this problem. To train these models we needed a lot of images. Data augmentation techniques are generally employed in standard deep learning problems. These range from scaling, rotating to inverting images which in this case would make the image unusable. However, this application needs a lot of data to train since we seek to regress both the position and pose from an image. Since collection and labelling of images accurately is time consuming, we developed a Blender add-on to simulate photorealistic images for training.

Once we know the relative pose and position of the camera with respect to the LED panel light, we need to identify the location of that light. We used optical camera communication (OCC) to facilitate light localization. This involves sending a unique ID corresponding to the light which is then received by the camera on the phone. The rolling shutter effect of CMOS cameras provides a temporal record of the different states a light has been in on the same image. This is due to each column of pixels in the image being exposed one at a time. We developed a simulation technique to generate images, which was then used to train a neural network to demodulate the ID from the received image. Though no modifications are required in the receiver, to send the unique ID using the

transmitter, we employ on off keying (OOK), which involves using an n-channel MOSFET as a switch to turn the LED panel light on or off depending on the encoded signal from an micro controller.

We have explored techniques to facilitate camera based VLP, but these work only when the user is under a light. When we are between lights, we propose using Wi-Fi fingerprinting to limit outage. We leverage existing open-source datasets for training and compare the model performance. However, the models trained on one building cannot be used for another since these models learn the relationship of a specific set of RSSs to the building and floor locations necessitating the expensive, time-consuming process of fingerprinting. Even when we consider the individual datasets producing these excellent results, painstaking optimization is required which precludes people from trying to implement indoor positioning quickly. Most of the input RSS vector is empty with redundant information and the static class labels used for buildings and floors make the models unusable on other buildings. We propose a machine learning-based framework that uses RSS values from the strongest access point (AP) signals and normalized output labels to combat this issue.

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background and Motivation

Indoor positioning is an interesting problem space where even the global positioning problem has found solutions achieving widespread adoption emerging as the clear winner. However, the indoor positioning problem has yet to arrive at a winner akin to GPS. This becomes important since the GPS signals suffer in indoor scenarios owing to walls and dense architecture [1]. This has become far more important especially with the advent of location-based services (LBS) and the internet of things (IoT). There are several new and novel applications of these services such as navigation, asset tracking, facility management, targeted advertising, check in and check out services, warehousing, virtual and augmented reality-based training. This along with the diverse applications of the internet of things which has grown to encompass everything ranging from smart refrigerators to sensor networks for indoor air quality monitoring.

The indoor positioning problem has been approached from several angles with widely accepted techniques such as Wi-Fi [2–5], radio frequency identification (RFID)[6, 7], Bluetooth [8–11], magnetic field strength [12, 13] and ultra-wideband (UWB) [14–17]. While the UWB, RFID and Bluetooth approaches provide excellent results of centimeter level localization, especially in the case of UWB they require expensive beacons and devices which can only be used for the purpose of localization. While these do have their own niche in the case of applications that span relatively small areas and require the highest possible level of accuracy such as drones [16], they are not scalable to cover large malls or universities that benefit from location based services. Most indoor spaces,

especially the spaces that seek LBSs, are already covered by a Wi-Fi network and they can leverage the magnetic field strength information which requires no additional infrastructure. Visible light-based positioning (VLP) techniques can be used to either replace or enhance the positioning accuracy of the aforementioned positioning techniques. There are several such approaches using photodetectors [18–21] or optical sensors [22–25] and they have much better accuracy since the area they span is very small it is referred to as an attocell [26]. If we apply the same logic of extra infrastructure required, the transmitters sending codes or enabling communication, necessitate expensive hardware but they do serve the purpose of faster communication nodes unhindered by radio frequency interference which plagues Wi-Fi based systems. This communication modality, termed Li-Fi was first introduced in 2011 by professor Harald Haas [26] and in the decade since has seen a lot of interest.

Thus, most of the IoT and LBS applications require decimeter level accuracy and providing this using GPS is not viable [1], moreover there is another issue in the case of buildings. This is the issue of figuring out the level a user is in, which is also impossible using GPS. With the advent of smartphones almost everyone has access to technologies employed for indoor positioning. Wi-Fi based techniques have seen widespread success and adoption mainly because of the value proposition of existing infrastructure providing additional functionality. There are approaches which bring together Wi-Fi and a slew of other sensors [27] along with other infrastructure free data modalities such as magnetic field strength, accelerometers and gyroscopes. Bringing such data forms together to fill a gap in the positioning process is also studied extensively under the pedestrian dead reckoning technique [28].

These bridging techniques are required mainly because of issues with Wi-Fi based techniques such as accuracy and latency. The overall positioning accuracy is quite coarse in most of the cases, more than a meter [29], while there are other improvements that can solve these issues and improve the accuracy, they bring problems of their own. In the case of channel state information (CSI), which can improve positioning accuracy of Wi-Fi, the calculation of these values requires driver level access, and this is not available in most cases and similarly in the case of round-trip time (RTT) newer standard routers are required, which is possible in the case of new networks if the cost is manageable but relegates all existing networks to lower accuracy.

The pursuit of visible light communication has persisted since the 19th century, with various attempts being made to repurpose light as the carrier wave for communication.

These efforts bore fruit in the beginning of the 21st century, with the use of light emitting diodes whose fast-switching speeds made it a viable mode of communication. The technology has now matured to the point where a fully networked communication system can be setup based on visible light communication, called Light Fidelity (Li-Fi) rivalling Wireless Fidelity (Wi-Fi). This has necessitated the sub-problems of latter, such as positioning, resource allocation, throughput improvement, be solved for the former as well.

The similarities of visible light communication and visible light positioning together to the Wi-Fi positioning proposition makes the former a good replacement or a supplement. The Li-Fi solution also provides high speed internet access apart from the positioning solution which becomes an additional advantage. Among the other extant techniques wi-fi is the only solution that can identify the user's level in a building, while in the case of all other techniques such as UWB and Bluetooth additional hardware is required. The same can be achieved in the case of visible light positioning with the additional communication functionality same as the wi-fi advantage. In the case of visible light positioning inexpensive microcontrollers can be used along with commercial off the shelf lights to send codes [30] which can be used to identify different levels in a building. The visible light positioning techniques are similar to the wi-fi solution in all of the advantageous aspects of the latter while bringing its own set of advantages on top of these.

The advantages of VLP counteract some of the disadvantages of wi-fi such as radio frequency interference which is one of the major reasons wi-fi becomes unstable for positioning over short distances, being influenced by occupancy and traffic in the access point being used for positioning. In the case of visible light positioning there are no such interference problems owing to the large visible light spectrum. There are concerns about the radio frequency radiation and its impact on small pets and babies in close range, which is countered by visible light positioning in that lights do not raise such concerns among consumers. The visible light positioning system also brings with it the line-of-sight (LOS) condition, which in most cases becomes a disadvantage owing to the limited coverage, but in this case, it prevents snooping on the network since walls block the light so a user will have to be inside the premises to connect to the network. This cannot be guaranteed in the wi-fi scenario which can be accessed past walls and even outside big buildings in the case of limited interference. The line of sight condition also gives much better accuracy in the case of visible light positioning with several studies achieving centimeter level accuracy [31].

The emergence of visible light communication (VLC) coincides with the development of better light emitting diodes (LEDs) egged on by the emergence of the commercial smart light phenomenon. LEDs act as the transmitter in the VLP schematic and their improvement has also served to improve commercial lights. The lifetime of an average LED has increased to nearly six years [32] which has made it commercially viable to see ubiquitous deployment of LEDs which can be used for visible light positioning. These lights have higher switching speeds and brightness compared to conventional lights which has established these as the clear winner in the market making it easier for the deployment of VLC in several commercial spaces using the lights installed avoiding the need for newer hardware.

The VLP techniques thus vary based on the receiver since the transmitter remains the same. The most common receivers are either photo detectors such as photo diodes and photo transistors or optical sensors such as the cameras. The photodetectors are much faster and accurate than the cameras, but they have to be retrofitted into the devices people use while most smartphones and computers come with cameras. The photodetectors have to be arranged in an array in the case of angle of arrival (AOA) and received signal strength (RSS) approaches or the system needs granular control over the transmitter in order to monitor minor parameter variations in the case of phase difference of arrival (PDOA) and time difference of arrival (TDOA) based systems. The major reason for cameras being an attractive alternative is the ubiquity of camera phones which does away with the need for an additional device as in the case of RFID or UWB based positioning systems. The craze for selfies also guarantees these cameras will get better over time making this an exciting prospect for VLP.

The use of camera phones brings with it a bevy of sensors which can track rotation, tilt and orientation with highly accurate results. The use of complementary metal oxide semiconductor (CMOS) sensors over charge coupled device (CCD) was cemented owing to higher speed and the rolling shutter effect which makes VLC possible on these images. The main challenge is in the form of the geometric problem that needs to be solved to track location, though additional sensors are available the additional data also has to be processed on board the smartphone since sending images to external servers can alarm users. The centimeter level real-time positioning problem becomes challenging when faced with the limited computational resource the problem has to be solved with.

1.2 Major Contributions

We aim to study techniques for accurate indoor localisation using smartphone cameras and commercial off the shelf transmitters. While photodetectors being incorporated into smartphones is certainly a possibility depending on future developments in VLC, CMOS sensor based VLP is the most viable option for solving the positioning problem with the devices and techniques currently at our disposal.

Indoor navigation for large areas such as malls and university campuses relies on indoor positioning. In the case of smartphones, front cameras have to be used to ensure the user can look at the screen, which tend to have a limited field of view compared to rear cameras. This coupled with the fact that lights are installed sparsely ensures that a light might not be fully visible within the field of view of smartphone cameras. While there are techniques that estimate one of the corners when three corners of a rectangular LED panel light are visible, none have reported results when only two corners are visible. Using a novel computer vision based technique, the corners were estimated when only two corners where visible leveraging 3D geometry and other sensors on smartphones to improve positioning accuracy. The proposed technique was also shown to perform better than state of the art positioning algorithms on images where the panel light was fully visible.

Large commercial spaces have a wide variety of transmitters with different shapes and sizes. Though images can be gathered after the lights are installed, simulation can allow for testing, which can in turn help in development of positioning algorithms and optimisation of transmitter placement. Thus a novel tree-based positioning technique was proposed leveraging simulated images. In the proposed technique,the simulated images were used to train machine learning models which in turn was used to test positioning accuracy on actual images. For want of a lightweight model without compromising on accuracy, the corners were extracted from simulated images and used to train models. Tree-based models were shown to perform better than the state of the art techniques.

Though the relative position of the receiver with respect to the transmitter can be identified quite easily, the localisation of the transmitter is more challenging. This can be done using ID matching through optical camera communication. The lack of a robust simulation technique to this end limits the ability to choose transmitters that allow for improved detection rates. A novel simulation technique was proposed and experimental validation shows the robustness of the technique to different transmitter and receiver

specifications. The simulated images were shown to be similar to the actual real world images and better than the state of the art simulation techniques. An improved machine learning model was trained to perform demodulation and bit error rates were shown to be better than conventional models.

The usage of commonly available infrastructure such as wi-fi for indoor localisation further improves positioning accuracy. A novel positioning technique leveraging existing open source datasets was proposed. Feature engineering was used to achieve high positioning accuracy using existing open source datasets and compared to the current state of the art positioning technique. The novel input and output normalisation schemes were shown to allow for the reuse of existing datasets on new buildings. The proposed technique managed to achieve higher floor and positioning accuracy on 11 publicly available datasets using lower computational effort and time.

1.3 Outline of the Thesis

The idea behind the thesis was introduced in the first chapter with the background of the problem and motivations being outlined. The major contributions and outline of the thesis were also delineated in this chapter.

A general problem description followed by different types of positioning techniques and systems were introduced in the second chapter. The system model was outlined to describe a generic camera based visible light positioning system along with the general positioning work flow. This distinction between different types of commonly used camera sensors and the reasons for choosing CMOS sensors was outlined. The role of the rolling shutter effect in the positioning process was highlighted. The range based and range free techniques commonly used for visible light positioning were also discussed. The computer vision problem at the center of the camera based VLP was explored following by deep learning based solutions. The use of sensor fusion was studied to bring together sensors on board smartphones and CMOS cameras.

A 3D geometry based technique was proposed and tested in the third chapter. The problem being solved was described with attention to the different coordinate systems and transformations between them involved in the same. The process of data collection using the experimental setup was described. The positioning technique was outlined with a focus on the corner detection and ordering process using the IMU data for heading along

with the image. The orientation, pose and positioning results for pictures where a rectangular LED was fully visible were reported and compared with results from the SOTA solution. The positioning results of the proposed technique for images where the rectangular LED was partially visible at two different heights were presented and discussed.

A machine learning technique based on simulated images was outlined in the fourth chapter. The need for a simulation technique was introduced and the issues with extant simulation techniques were explained. The proposed flow of the simulation and indoor positioning pipeline was detailed. The experimental setup used for data collection for training and setting was delineated. The choice of model was discussed through reported results from the simulated and experimental data. The mean 3D positioning error was compared with a SOTA technique. The ability of the trained model to generalise learned information beyond the data available in the training dataset was tested for different heights.

A OCC technique for indoor positioning was proposed and tested in chapter five. The problems with current techniques were explained and the proposed technique was explained. The experimental setup used for collecting a testing dataset was detailed. The current conventional and proposed demodulation techniques were outlined. The simulated images were compared to experimental images using the discrete Fréchet distance. The influence of different transmitter parameters such as shape, size and luminous intensity on the proposed simulation technique were tested. The effect of image noise on decoding techniques was tested. The effect of receiver parameters such as exposure time, aperture, resolution on the proposed simulation technique were reported. Different modulation and encoding techniques were compared using bit error rate from the proposed and conventional demodulation techniques.

A wi-fi based indoor positioning technique was proposed and tested in the sixth chapter. The need for the proposed technique and failings of extant wi-fi fingerprinting based indoor positioning techniques were highlighted. The proposed technique with its constituent parts of transmitter localisation, input and output normalisation were outlined. The effect of the transmitter localisation technique, data normalisation, model selection and feature extraction on floor and positioning accuracy were studied. The feature importances were analysed to ensure the results were improved because the models learnt the relevant input output relationships. The generalisability of the proposed technique was tested using eleven publicly available open source wi-fi fingerprinting datasets.

The conclusions drawn from the work presented along with avenues for future work were discussed in the seventh chapter.

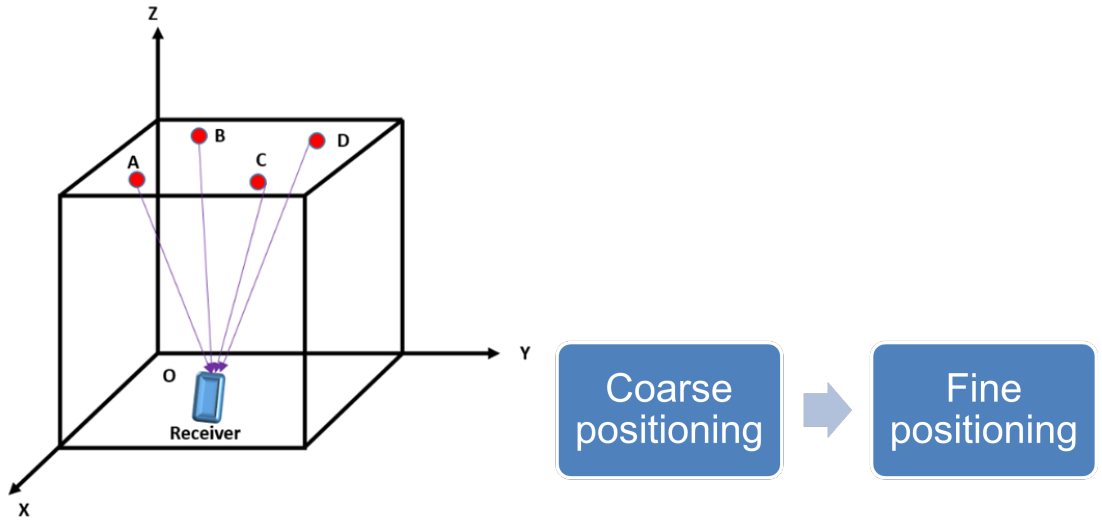
Chapter 2

Literature Review

2.1 Introduction to visible light positioning

The inability of the global positioning system to provide accurate positioning results for the indoor positioning problem has led to the adoption of several other techniques such as wi-fi, magnetic field strength, Bluetooth, visible light positioning and UWB. Among these the techniques that can leverage existing infrastructure stand out owing to lower cost of deployment, maintenance and over the top functionality. The wi-fi positioning technique has seen significant advances but has recently seen several setbacks including the android ten wi-fi scan throttling decision, which has left extant systems using wi-fi positioning technique with no viable option. The visible light positioning system is clearly the frontrunner now as it brings several of wi-fi advantages along with a few improvements of its own. This technology is in its incipient stage and will only get better as a common IEEE standard is established and both transmitters and receivers improve with widespread adoption.

The visible light communication paradigm is similar to radio frequency communication in the modulation, multiplexing and coding requirements but varies in the part of the electromagnetic spectrum being used as the carrier wave. The VLC system uses carrier waves from the 400 to 790 terra hertz frequency as opposed to RF which uses the twenty thousand to three hundred billion hertz frequency range. The much larger bandwidth in the case of VLC and the lack of regulation means no interference and no restriction on the techniques that can be used. This interestingly opens the flood gates to newer techniques which could not be applied to earlier systems. One such area that has seen



(a) Representation (b) Steps

FIGURE 2.1: A generic VLP system.

major improvements owing to the nature of the VLC system is visible light positioning. VLP much like wi-fi based positioning system serves as an additional feature apart from the exciting high speed communication possibilities. In this section the VLP based techniques will be reviewed with an emphasis on camera-based systems, image processing and sensor fusion.

2.1.1 System Model

The basic representation of the problem at hand from the end user point of view is outlined in Fig. 2.1(a), where the receiver is represented using a smartphone since it could be retrofitted with photodetector or use an inbuilt camera as the processing input device. The global coordinate system is delineated by the x, y and z axis with the origin at o. The points A through D are the transmitters, which will be light emitting diodes in most cases, which could be in range of the receiver. In the case of a photodetector array being the receiver multiple lights could be in range which paves the way for range-based detection techniques and the process of using multiple transmitters to pinpoint the location of a receiver is employed.

The problem can be broken down to two constituent parts as shown in Fig. 2.1(b), where finding the position and orientation of the receiver with respect to the transmitter is referred to as the fine positioning problem in this thesis and finding the position of the

transmitter in the global coordinate system is referred to as the coarse positioning or transmitter localisation problem. The coarse location refers to the building, floor, room and position where the transmitter is located. These parts are then put together to find the position and orientation of the receiver in the global coordinate system. The Euclidean distance between the actual position of the receiver, which is the ground truth, and the estimated position, which is the output of the VLP system, is the accuracy measure of the system.

2.1.2 Positioning Workflow

The positioning workflow leverages the simplex communication modality incorporates a code being sent to the receiver through modulation schemes on the transmitter side. The location code is modulated into the lighting source through an appropriate modulation technique. For a technique that works on any device the simplest of such techniques called on off keying (OOK) could be used. As the name suggests this modulation scheme employs two levels to denote the two possible bits that could be sent at any point in time high for one and low for zero bit. This is the first step of positioning which is then followed by processing the data to recover the location information being sent by the transmitter. This can be done using image processing or signal processing depending on the receiver being used for detection. The final step involves estimation of the receiver location and orientation in the global coordinate system.

The initial step at the transmitter side is where the combination of the signal of interest, which is the location tag in this case, and the carrier wave in the visible light part of the electromagnetic spectrum are combined through a process called modulation. The data is first encoded to a different format both for security and to improve ease of detection. These also come with error detection and correction codes which could range from something as simple as a parity or cyclic redundancy check to complicated coding techniques such as turbo codes. The location tag is thus combined with both the error detection and correction code along with a header to show where one string ends and the next begins. The limitation of the receiver in the case of a camera to thousands of hertz necessitates the multiplexing of data to incorporate as much data as possible in the same channel, which could be time division or frequency division-based techniques. Once the data is

encoded modulating the alternating signal into the direct current-based LED is done using a bias tee. This data is then transmitted over the wireless optical channel (WOC) and received by the receiver of choice at the user end.

The receiver could either be a photodetector or a camera which are both capable of identifying the location tag being transmitted by the LED source but achieve their results through different processes. The photodetector converts the transmitted signal directly into an electrical current signal which is much more sensitive to the intensity than an average camera and much faster making it the automatic choice for visible light communication systems. These photodetectors when combined to form an array or when combined with a camera to form a hybrid positioning system provide a complete system of features to enable communication and positioning. Once the received signal is demodulated and decoded the location tag can be used to identify the location of the transmitter in the global coordinate system.

If the receiver used is a camera two types of responses can be expected depending on the type of camera being employed. The two major types of cameras are charge coupled device (CCD) sensors or complementary metal oxide semiconductor (CMOS) sensors. The former is generally unusable in the visible light communication and positioning problem space owing to its low speed and the general philosophy of operation. The CCD sensor converts light to electrons, same as a CMOS sensor, but is an analog sensor. The CCD sensors are generally costlier and more power consuming but produce higher quality lower noise photographs. The advantages of CCD sensors in the case of photography become issues when dealing with VLP problems. The higher quality is provided by combining information from the environment at the same time instant, which removes distortion caused by moving subjects and reduces the frame rate of these cameras owing to the higher amount of time required for processing all individual pixels at the same instant. The CMOS sensor is a digital sensor which brings with it advantages of being the low cost and low power alternative and a specific defect in the case of photograph which makes it invaluable in the VLP and VLC problem space. The rolling shutter effect which causes distortion in moving subject improves the refresh rate to much higher levels. This is caused by the CMOS sensors collecting and scanning individual columns continuously and combining the image as such. Hence in an image from a CMOS sensor the elements in the left of an image are captured earlier than the elements in the right part of the image depending on the scanning direction as shown in Fig. 2.2(a). This form of continuous capture of sections of the image works for the VLC problem where the light

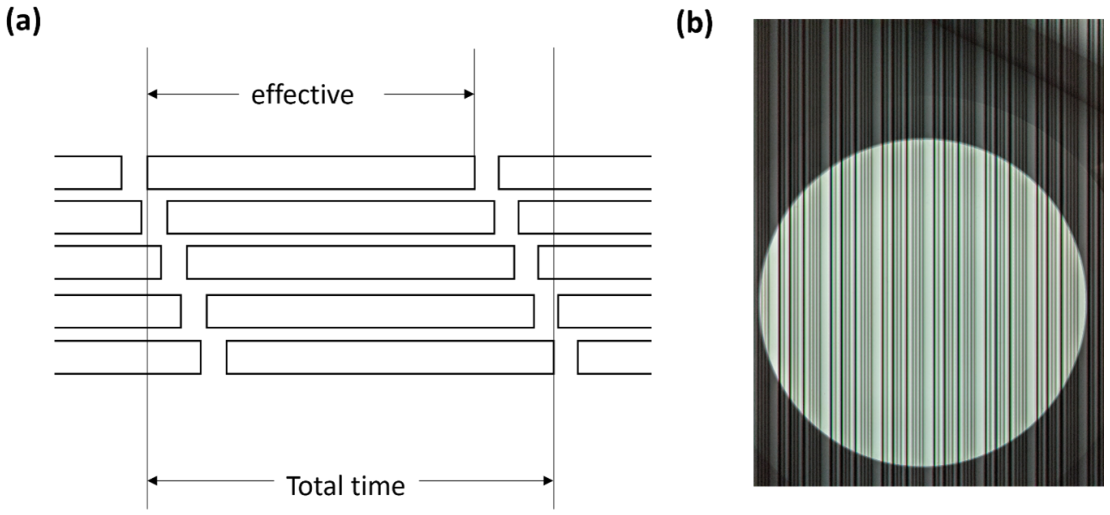


FIGURE 2.2: CMOS sensor rolling shutter representation.

remains stationary, but it is turned on and off at a high frequency. The different states a light goes through are documented clearly in an image from a CMOS sensor with the data being captured from left to right or from top to bottom moving from the oldest to newest. Such an image is shown in Fig. 2.2(b), where the white bands are where the transmitter is sending bit one and the bark bands when bit zero is being sent. The duration for which the lights are turned on and off are also captured in these images with the width of the dark and bright bands being a representation of the same. Thus, CMOS sensors provide the unique advantage of the individual bands being obtained much faster than the overall image as shown in the Fig. 2.2(a), where the individual first block represents one of the image columns which is captured at the effective time of capture much lower than the total time of capture for all columns in the image put together.

These advantages make the CMOS sensor tailor made for the VLC problem space and the fact that these low-cost sensors are in all camera phones will only make them better with time. The major limitation here is the effective time for compiling one column of the image if the compile time is greater than the time period for which one bit is on or off in the modulation scheme used at the transmitter the system will no longer be capable of receiving information. So, the frequency of modulation at the transmitter side must be lower than the line scan rate to ensure proper functioning of camera-based systems. However, photodetectors have much higher frequency tolerance and are hence preferred for visible light communication problems with a focus on data rate improvement. Even in the case of VLP complicated modulation schemes, which tend to provide additional functionality, benefit from higher frequency of operation and this can be achieved only

with photodetectors. The location tag obtained from the receiver is then combined with features from the image, in the case of CMOS sensors, or the signal, in the case of photodetectors, to arrive at the final leg of positioning to ascertain the position and orientation of the receiver with respect to the transmitter. This information is combined with the position information of the transmitter in the GCS from the location tag to obtain the position and orientation of the receiver in the GCS.

2.1.3 Range based techniques

These could be used in the case of time of arrival (TOA) or received signal strength (RSS) being used as the metrics for identification. The receiver can identify the amount of time taken by the signal to reach the sensor from the transmitter and can hence be used to find out the distance from the transmitter since the velocity of the wave and time taken are known. This can also be mapped to the signal strength which generally decreases with increase in distance from the transmitter which can be used to estimate receiver location. The technique however requires at least three transmitters to be in range and

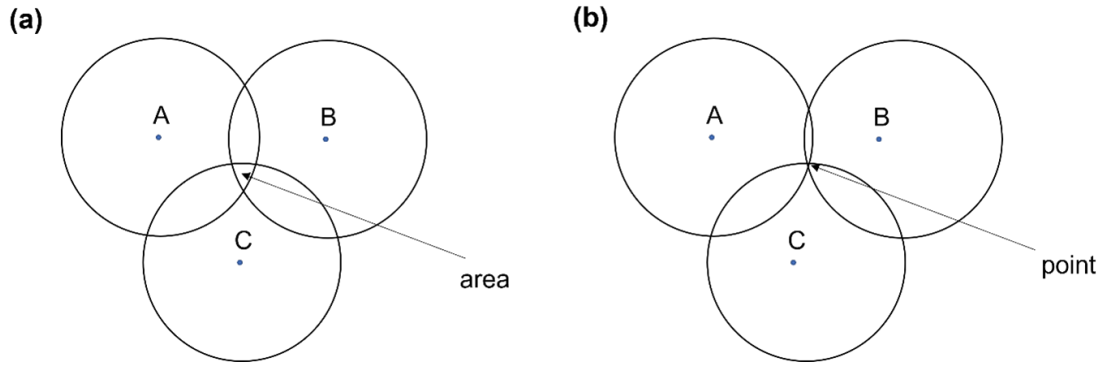


FIGURE 2.3: Trilateration representation.

is hence called trilateration. The general representation of trilateration in this system model is shown in Fig. 2.3, with the transmitters being represented by the alphabets A, B and C. The transmitters could give pinpoint accuracy in two dimensions such that the information from the three transmitters coincide at one single point on the GCS as shown in Fig. 2.3(b) or narrow down to an area as shown in Fig. 2.3(a). These techniques have been explored extensively for both photodetector and camera-based systems[31], [33]. The RSS based problems lead to the solving of a system of simultaneous equations as follows.

$$(x - x_1)^2 + (y - y_1)^2 = d_1^2 \tag{2.1}$$

$$(x - x_2)^2 + (y - y_2)^2 = d_2^2 \tag{2.2}$$

$$(x - x_3)^2 + (y - y_3)^2 = d_3^2 \tag{2.3}$$

The system of equations represents the x, and y coordinates of the three points, which are the locations of the transmitters in two dimensions, and the corresponding distance marked by their subscript. In the case of three-dimensional positioning, the circles would be replaced by spheres marking coverage and the system of simultaneous equations to be solved will also incorporate the z coordinate. The RSS based positioning can be done with photodetector as receivers or with camera-based solutions [23], [34–36]. RSS provides for a reliable measure which is also easy to access which is why most VLP systems especially photodetector-based systems tend to default to signal strength. RSS manages to provide three dimensional accuracy of less than three cm using either carrier allocation without any auxiliary devices and three light emitting diodes employing frequency division multiplexing [36] and less than three cm two dimensional accuracy in a larger area by using trilateration [35].

Time of arrival (TOA) and phase of arrival (POA) are metrics employed for range-based positioning. TOA calculates the position of the receiver based on the product of the time taken by the light to reach the receiver from the transmitter and the speed of light which is known for the conditions[37]. The POA approach uses the phase of the carrier wave at the transmitter to check the shift in phase at the receiver. This is subsequently used to find the distance from the transmitter since it is the product of wavelength and the shift in phase. These geometric techniques require synchronisation between the transmitter and receiver and high accuracy on the receiver side to ensure proper functioning. These schemes are generally avoided for the strict synchronisation and accuracy requirement and the lack of accuracy and positioning improvements to justify the additional requirements.

Time difference of arrival (TDOA) and phase difference of arrival (PDOA) are techniques employed to counteract these requirements. They use differences in phase and time instead of the value by itself thus reducing some of the computational requirement. While the TDOA based techniques have managed to achieve sub cm two dimensional localisation accuracy [38] and sub mm three dimensional accuracy through simulation in the case of photodetector [39], they still have difficult to meet requirements with respect to accuracy and resolution for time synchronisation and measurement which has put off people from achieving such high accuracy metrics in experimental demonstrations. The

PDOA systems generally face the same issues but when combined with RSS these systems have managed to achieve practical accuracy using photodetectors in the range of ten cm [20].

The angle of arrival (AOA) is the major issue as far as visible light positioning is concerned since determines the direction of the transmitter with respect to the receiver, without which however accurate the distance from the transmitter is determined to be the whole operation could fall apart. To perform this important aspect of the positioning process both receivers struggle with both needing at least two of each to perform accurate triangulation, since even a small deviation can be exaggerated when extended further from the transmitter. AOA does away with many requirements such as receiver transmitter synchronisation, high receiver resolution, modelling of power or multi path effects. While simulation results show high two dimensional positioning accuracy of less than fifteen cm [40, 41] the experimental AOA triangulation based approaches bring together either inertial sensor [42] or an accelerometer [43] by itself to achieve sub thirty cm two dimensional accuracy. While these range-based techniques provide good accuracy they are tailor made for photodetectors with most metrics being impossible to calculate using imprecise CMOS sensors.

2.1.4 Range-free techniques

While triangulation and trilateration seem like straightforward solutions to the problem at hand, they require accurate readings which rules out most inexpensive sensors and increases computational complexity of the problem. The slew of techniques used to solve this VLP problem range from fingerprinting to vision-based approaches. With these techniques allowing for the use of lower resolution sensors most studies choose fingerprinting with RSS as the metric of choice to be monitored or modelled [31]. The fingerprinting process involves creating a database of values observed at different locations and estimating the position of the receiver based on the measured value's proximity to the value in the database corresponding to the location in question. Several studies have managed to achieve sub ten cm 2D positioning accuracy through photodetectors [44, 45] and using CMOS sensors [46] but this still falls within the specialities of the photodetector owing to modelling and simulation bestowed by the additional accuracy of such sensors.

The CMOS sensors achieve similar localisation accuracy through the vision-based techniques such as transformation and scene recognition. They can also use probabilistic

techniques for fusion with other sensor information which is important to bring together the positioning accuracy and information missing in the CMOS sensors. The triangulation and trilateration problems fail in their scalability since knowing where a receiver is in a small space with few lights does not bring the issue of differentiating between transmitters to the fore. If there are multiple lights knowing where the signal is coming from becomes difficult rendering most of the metrics discussed in the range-based techniques useless.

The fingerprinting problem also becomes much more challenging to solve with multiple lights, these can all be solved using the positioning workflow detailed earlier also known as ID matching. This performs the task of differentiating between multiple transmitters and leverages the advantages of CMOS sensors. With ID matching for CMOS sensors the light source transmits a signal which is decoded and compared against a database of known codes and corresponding locations in the GCS to identify the light source and further positioning techniques such as image processing-based solutions or machine learning algorithms which bring together the probabilistic techniques and vision-based techniques can be employed to identify the position and orientation in the GCS accurately. Most of current studies discussed here fail the transmitter differentiation test thus failing the important criteria of scalability. This vision-based problem along with deep learning using CMOS sensors and sensor fusion are studied in detail in the following sections.

2.2 Image Processing Problem

The CMOS sensor brings certain unique advantages to the table most important of which is widespread commercial adoption in the form of camera phones, in tablets and in most mobile devices. The inexpensive, low-power nature of these sensors and scope for improvement even without the commercial adoption of VLC, given cameras in phones and mobile devices has been improving year on year solidifies the case for solving the VLP problem using CMOS sensors. Given that these sensors suffer when measuring accurate metrics and when using these metrics for fingerprinting, the best alternative is solving it as an image processing problem. The problem formulation at its core is one of pose estimation, which requires a deeper understanding of the image capturing process.

The image of the LED source is captured by the CMOS sensor and using the image the position of the camera is to be obtained. The coordinate system with the center of

the CMOS sensor at its origin can be called the camera coordinate system (CCS). The relationship between the image coordinates and the CCS can be obtained based on the extrinsic and intrinsic parameters of the camera being used. This extends to finding the real-world coordinates of the transmitter corresponding to the coordinates of the transmitter in the image plane. These transmitters come in a variety of shapes and sizes ranging from point sources of light to large rectangular panels of area lights. Depending on the shape of the transmitter the ability of the algorithm to extract features of interest vary, since a rectangular light will have four points at the corners and edge lines connecting these corner points can be used as features of interest, similarly in the case of a circular light the center of the circle and its diameter can be used as the features of interest. Once mapping between the image plane and the CCS is obtained the problem is further simplified into one of finding the translation and rotation of the receiver with respect to the transmitter. The cluster of features indicating the transmitter can be identified as the object of interest and its translation from the CCS to GCS can be represented by the equation as follows

$$P^G = R_C^G . P^C + t_C^G \tag{2.4}$$

The coordinates of the points representing the transmitter in GCS and CCS are be represented by P^G and P^C respectively. The rotation translation matrices from the CCS to GCS are represented by R_C^G and t_C^G respectively. Thus, we can see the problem reduces into the identification of the rotation and translation matrices which can be achieved through transformation techniques. In certain cases, the identification of the rotation matrix might be impossible without the use of auxiliary sensors such as in the case of a circular of square transmitter where the projection of such shapes on the image plane will be uniform on all directions making it impossible to detect tilt and hence pose reliably. There are several studies where the features mapped to transmitters are obtained from multiple transmitters [25, 47–49] in an image which makes the resulting solution to the problem a lot more reliable, in the order of less than ten cm accuracy.

While such techniques utilizing multiple transmitters, at least three such transmitters then suffer a drop in accuracy when those many lights are not in the field of view of the camera. In commercial and residential spaces, the focus as far lighting is concerned is to reduce the cost and the number of lights used for illuminating the area. Hence, the fewest number of lights required will be spaced out far away from one another. This combined with the fact that most of these applications will be tested on camera phones of which the front facing camera is used since the user will also need to see the screen and simultaneously

point the camera at the lights which are in the same direction as the user's head ensure that the field of view of the camera being used is not big enough to cover more than a couple of lights in the image at any given time.

The technique should therefore use a single light as the basic unit of testing and the positioning accuracy obtained from the same should be used as the benchmark for comparison. The studies discussing such results with single LED are few and far between. A study [22] proposes the use of VLC assisted perspective four line algorithm to perform camera-based positioning in a cell with a single rectangular LED where the position and dimension of the transmitter are known. They produce simulation results where the accuracy never falls below fifteen cm and in experimental testing are able to show three cm accuracy and four-degree orientation accuracy.

This however assumes that all four corners and edges of the transmitter will be visible, and the camera being used will always have the intrinsic and extrinsic parameters from a calibration. Another study[30] with circular lights deals with this problem by combining circle projection with a singular value decomposition based algorithm to solve the positioning problem. In order to find the orientation of the receiver with respect to the transmitter the azimuth or the deviation of the receiver from the magnetic north is estimated by using a marker on the light at a predefined position. While these are interesting solutions to the problem, they all have the common problem of having to know the shape and size of the LEDs.

When the shape and size of the LED changes the techniques fail because the complex geometric projection equations work only for the defined shape and size. This coupled with the fact that most spaces have a combination of different shapes and sizes of lights even in the same room ensures that these techniques breakdown quickly in real-world situations. The solution should be capable of working for multiple light shapes and multiple sensors since the different cameras will have different parameters which might not fit the requirements of the predefine system or should be easy to retrofit to the problem space. The machine learning-based solutions are capable of satisfying these requirements with sufficient data and computation capability, these possibilities will be explored further.

2.3 Deep learning solutions

Over the past decade machine vision has witnessed groundbreaking improvements with the advent of deep learning. Though machine learning-based solutions could be used to represent any of the techniques ranging from basic clustering techniques to convolutional neural networks (CNNs) the problem space requires the more complicated deep learning techniques to solve. While the specific problem of camera-based VLP for indoor positioning has yet to be broached the broader problem of identifying camera pose from monocular images using deep learning has been studied extensively [50]. This type of problem is referred to as a pose estimation problem or a camera relocalization problem.

The nature of the problem necessitates a deep learning-based solution, which can be broken down into two major types, one where the features of interest are identified and extracted manually and the other where the model is allowed to extract the features through end-to-end convolutional structures [51]. The structure of the proposed network performs classification to pinpoint the location in a discretized initial step with the subsequent regression problem becoming much easier to solve. We have seen that end-to-end convolutional networks are capable of learning features and generalizing the same to other classification tasks [52–55] which is an important facet of the problem space where a new requirement for shape of the light or specification of the receiver or quality of the lighting could require adjustment to the trained model and a classification step guarantees useability with a small dataset generated to account for the specific changes of interest.

The idea of transfer learning becomes important owing to the complicated problem space, a regression problem from images using CNN is extremely complicated due the wide range of values each output can take and the seemingly infinite variations ranging from the environmental lighting to light shape that need to be accounted for in this problem space. With the addition of a classification step much needed generalization is guaranteed which can then be retrained based on the newer variations of interest. Though similar regression problems of two dimensional joint position estimation have achieved excellent positioning accuracy [56], they do not have the same wide range of movement possible in the VLP problem space with most of the joint positions constrained to the small movements on the image plane. The Pose Net architecture [51] is a modified GoogLeNet network [57] where the final prediction layer has been replaced with four fully connected layers to perform the regression task at hand. The general structure of these type of networks is shown in Fig. 2.4, where encoder learns features from the image and compresses

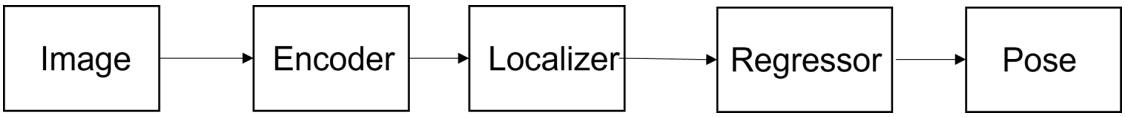


FIGURE 2.4: General pose estimation network structure.

them into a dense array of important features which is then followed by a fully connected layer called the localizer which maps the features to a feature vector. This is followed by the regressor which is a set of two separate dense layers which give the output pose in a quaternion form comprising of both the position and rotation. Though the pretrained network was used the proposed system fails in replicating the success of [57] to unseen data. This failure in generalization to newer defeats the purpose of using transfer learning and the high computational cost of the deep network with a hundred and forty-four layers.

The network also produces results far behind the state of the art in manual feature extraction based techniques [58, 59] thus fomenting interest in the network design for the camera relocalisation problem. Some newer networks have attempted modifications on the original structure with varying degrees of success. A network structure named LSTM-Pose[60] which modified the localizer to include four long short term memory networks (LSTM) which improved performance over the existing network. While better results were achieved with the network structure named VLocNet[61] which surpassed original structure based accuracy for indoor scenes by incorporating the ResNet50[62] architecture for the encoder and incorporating two images at subsequent time instances into the input thus leading the encoder to learn features from both these images in separate branches. While this advanced positioning algorithm has seen quite a bit of interest in recent years it still has a wide variety of input variations that the problem needs to solve for and the specific problem space of indoor positioning using lights would simplify it to a smaller subspace that has yet to see significant improvements owing to the niche nature of the problem.

The results here are obtained by training on publicly availably hand labelled image datasets such as ImageNet[63] which contains fourteen million images and Places [64] which contains seven million labelled images. These datasets have become benchmark for testing results of network structures, while there are indoor datasets the lack of a dataset to suit our particular problem space is preventing optimization and even recognition of the problem among those working on the pose estimation problem. The impact the ImageNet dataset has had on the development of object classification network cannot be ignored.

Thus, the creation of a dataset is integral for widespread recognition and common benchmarking of results among existing solutions.

There are two ways to create large datasets of images for classification and regression, where one involves collecting the images and labelling them manually the other involves simulation. There is precedence for simulation having produced datasets or dataset augmentation techniques. In the case of lidar point cloud generation for semantic segmentation in the autonomous driving application a combination of game mechanics from grand theft auto video game and ray casting produced a viable dataset [65] in combination with the kitti dataset[66]. There is a model based object detection dataset called the shape net data set [67] which contains models of common items like chairs with several variations introduced which was used to generate images and train a network to detect real-world objects like chairs without ever having shown the network actual images of the object in question.

The ability to simulate data also aids in improving the performance of deep learning networks by pretraining on simulated data as in the case of photodetector based indoor positioning scheme [20], where a combination of RSS and PDOA are modelled and the network is pretrained on this data which helps the network learn which a much smaller dataset of real-world data. Hence, the importance of a dataset that covers all possible variations of interest is apparent in the pose estimation problem space. The objective is to simplify real world data collection through videos and using structure from motion to label large number of images automatically [51]. Though this will create a dataset of several thousand images very quickly it still cannot cover all possible variations since a variation in light shape would require actual equipment cost and labor of installing the new light which quickly becomes infeasible for a wide representation of all possible problem cases which can be achieved through simulation.

2.4 Sensor Fusion

The indoor visible light positioning problem with a focus on CMOS sensor-based implementation faces several pitfalls with regards to the feasibility both for single cell implementation of orientation detection from a single camera and large-scale testing for multiple lights since the establishment of a baseline for testing would be expensive and introduce a significant barrier to entry for researchers exploring the problem. The need

for fusion with other data modalities is apparent when looking at common pitfalls of vision-based positioning techniques. The VLP studies with single LED have to either use other sensors on board for tracking orientation [22] or come up with novel techniques incorporating external markers for orientation tracking [30] .

Hence fusion of inertial and image sensors is a popular combination to offset the lack of directionality and depth information from a single CMOS sensor [31, 47]. Further, the CMOS sensor based VLP suffers from the same problem of differentiating between multiple transmitters. While ID matching involving a coded light module and establishment of a location ID based database for matching a transmitter's code to its location could be possible having all lights be modified purely for the sake of positioning is on the same scale as new infrastructure-based positioning schemes involving UWB or Bluetooth techniques. Hence, until widespread adoption of VLC by consumer device manufacturers the feasibility of VLP system has to be in improving and working in tandem with other extant positioning techniques such as wi-fi or pedestrian dead reckoning based systems.

The implementation of sensor fusion for camera-based VLP is thus observed in two stages for intra-cell orientation and depth tracking with another camera, or inertial sensors and inter-cell position identification through fusion with ID matching, photodetectors for VLC, magnetic field strength or wi-fi. Of these choices for inter-cell positioning wi-fi is the preferred option owing to the advantage of being the only other technique capable of highly accurate level and room identification in buildings [68]. Several wi-fi based studies incorporate other data from a wide variety of sensors including magnetometers [69–71] and pedestrian dead reckoning [72]. While there are techniques that employ such sensor fusion effectively the use of deep learning has also seen a steady increase in the wi-fi indoor positioning problem space over the past decade[29].

One of the major problems with indoor positioning techniques in general is the wide variety of input structures and metrics along with the lack of a consensus on the best metric for tracking the performance of a technique. In the case of wi-fi data one of the popular input metrics is received signal strength (RSS), which is by far the most popular choice especially for fingerprinting [29], where the structure in which the data is organized varies greatly depending on the number of access points (APs) being surveyed which in turn is dependent on the overall area being surveyed. The APs could be organized as an image simply for leveraging image based techniques such as deep CNNs [68] or based on previous steps taken to track orientation with the wi-fi data alone [73]. While there are other techniques such as channel state information (CSI) or round-trip

time (RTT) they have not been studied as extensively as RSS based techniques owing to the inherent complexity in data extraction for other techniques.

With the addition of standard datasets [74, 75] though this situation has improved greatly in the indoor positioning space due to clear splits in training and testing data ensuring proper benchmarking and straightforward comparison. The creation of datasets however brings with it the problem of data collection which is solved by employing crowdsourcing [71, 75]. The dataset thus created encouraged the use of several complex deep learning architectures. However, the best accuracy was achieved by simple models such as multi-layer perceptron (MLP) or auto encoders (AE) to achieve sub meter positioning accuracy with wi-fi RSS data alone [76, 77]. These results indicate that wi-fi based positioning is here to stay and combining the CMOS sensor-based positioning system with extant wi-fi networks through deep learning by leveraging existing standard datasets is a problem of interest.

Chapter 3

Computer vision based fine indoor positioning technique

3.1 Introduction

The VLP techniques use either photo detectors (PDs) or cameras as receivers. While photo detectors are capable of high data rate visible light communication (VLC) when compared to cameras, they are worse than cameras at localisation. The cheap CMOS sensors available on smartphones are an obvious choice for techniques that seek practical implementation of VLP in extant spaces. The location of the transmitter must be identified after which the location of the receiver with respect to the transmitter can be identified. The transmitter identification problem in this case is an indoor positioning problem using smartphones, which has been explored for several decades. Radio fingerprinting techniques have been shown to provide accurate floor and building detection across several datasets[78, 79], and magnetic field strength has been shown to achieve sub meter accuracy[27, 80] which is close enough to identify the transmitter location owing to the sparse distribution of indoor lights. VLC has also been employed with cameras for light identification[22, 30, 81] by encoding unique codes in lights acting as beacons. Owing to the many well established options for transmitter identification, this chapter will explore the relative pose and position identification assuming the transmitter location is known.

The camera-based single LED positioning problem has been explored extensively. In the case of circular panel lights, circle projection was used to estimate position and a

red marker on the light was used to identify orientation in [30] while geometric features from consecutive frames were used to improve positioning accuracy in [81] and plane intersection with line scheme was used to facilitate positioning in [82] . Ellipse fitting was used for positioning and projective geometry was using to calibrate an IMU in [83]. In the case of rectangular panel lights, corners are widely used for positioning. Corners were used with the perspective n point (PnP) approach in [84], with random forest regression in [85], with the perspective n line (PnL) approach in [22]. Since at least three known point correspondences are needed in these approaches, they have yet to be applied for partially visible lights in images.

The proposed technique seeks to address this gap by using two known points to estimate two more points on parallel sides of the rectangular panel light. These estimated points allow the use of the PnL approach. While the fusion of inertial sensor information for the identification of azimuthal angle is well explored with deep learning[86] and graph optimisation[87], it produces erroneous results due to uncalibrated IMU on smartphones. The proposed technique seeks to identify the general direction in which the phone is pointing when the image is taken reducing the accuracy requirement of the IMU.

The main contributions of this work are listed as follows

- A novel single LED VLP technique that outperforms the current state of the art technique.
- The first VLP technique to detect pose and position when only two corners of a four corner LED are visible.

3.2 Methodology

3.2.1 Problem Description

The layout of a camera-based single LED VLP problem is shown in Fig. 3.1(a), where a square LED panel with its corners marked $P1$ to $P4$ is in the ceiling of a room. The axes of the world coordinate system (WCS) are marked with subscript W and the axes of the phone coordinate system (PCS) are marked with subscript p. The axes marked N, E and U point towards the magnetic north, geographical east and up directions respectively. The magnetometer on the phone will provide the orientation of N with respect to the

PCS Y_p axis. The height of the camera from the transmitter is marked h_t. An image of an LED panel from the camera on the phone is shown in Fig. 3.1(b), where the pixel coordinate system (PiCS) with its origin in the top left of the image.

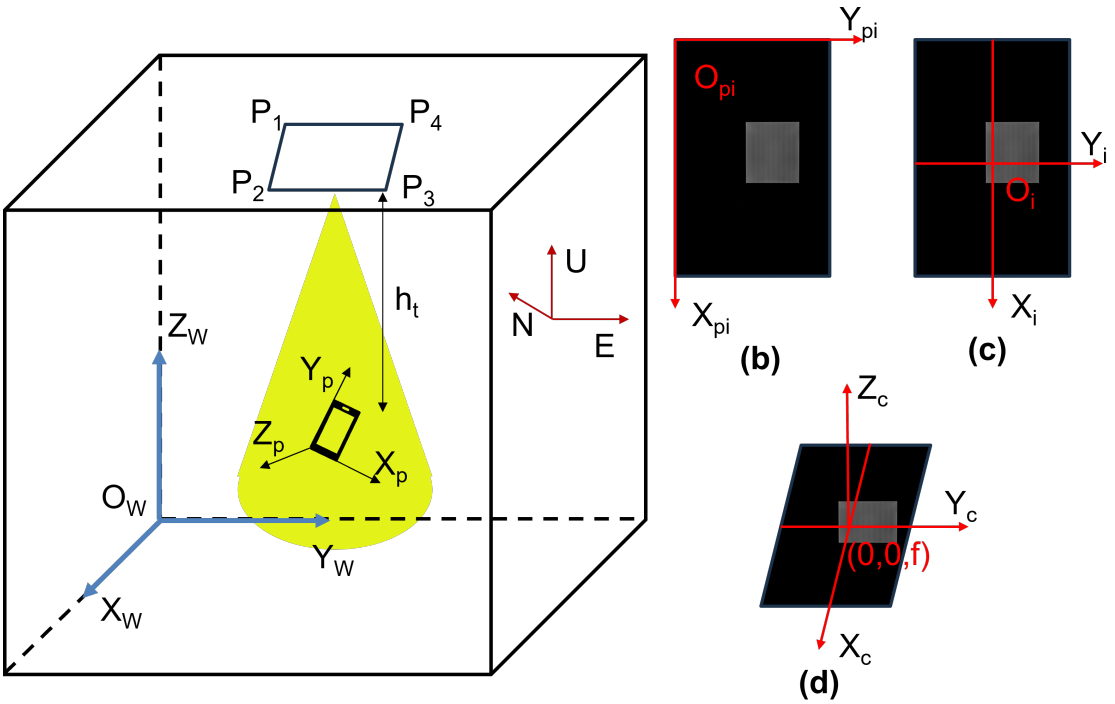


FIGURE 3.1: Coordinate system layout.

The image coordinate system (ICS) shown in Fig. 3.1(c), moves the origin O_i to the centre of the image and converts the units from pixels to meters. The centre of the image is subtracted from the coordinate of a point in PiCS and multiplied by the size of a pixel in meters to get corresponding ICS coordinates. The camera coordinate system (CCS) places the image plane in 3D space at focal length f from the CMOS sensor as shown in Fig. 3.1(d). The rotation and translation of CCS with respect to the WCS provides the pose and position of the receiver with respect to the transmitter. There are two parts to the positioning problem, identifying the location of the light and estimating the relative position of the receiver. This chapter will focus on the second part while the light location and its orientation with respect to the magnetic north are assumed to be known accurately.

3.2.2 Experimental Setup

Images were captured on a Redmi Note 9 Pro smartphone placed on a tripod as shown in Fig. 3.2(a), with the position being tracked using a 1.2m X 1.2m grid of tape on the

ground, where each cell was 40 cm apart, and the pose being tracked using an inclinometer, for pitch and roll, and a compass for azimuth. Five images were captured at each location on the grid with the phone facing the door and away from it, along with IMU readings using sensors on board the smartphone. One set of images was captured with different poses always ensuring all four corners of the light were within the FoV of the camera and another set was captured such that only two of the LED corners were in the FoV.

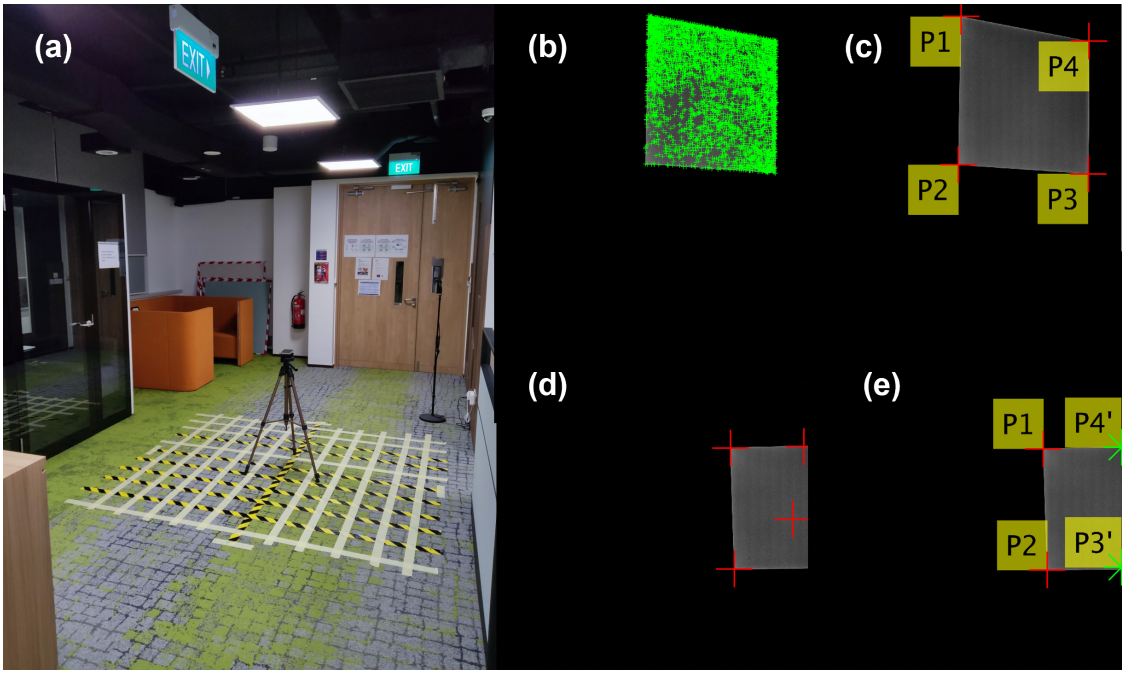


FIGURE 3.2: (a) Experimental Setup (b) all detected corners (c) strongest 4 labelled corners (d) incorrect occluded light corners (e) corrected labelled corners.

The images were also captured at four different heights from the transmitter ranging from 120cm to 156cm in 12cm increments. Thus 640 images comprising of 5 images with different poses at 16 grid locations facing two different general directions at four heights was captured for the complete light dataset. The general direction was found to have minimal impact on the positioning accuracy, so the partial light dataset was captured in only one direction with 120 images comprising of 5 images with different poses at only the outer 12 grid locations at 120cm and 156cm from the light. These images, the ground truth pose and position along with the known light location and orientation were used for localization.

3.2.3 Corner Detection and Ordering

Structure based VLP techniques leverage known points in WCS and CCS. Since the images were captured at 68 microsecond exposure time with 100 ISO only the light was visible with a black background. This was done to facilitate VLC where the exposure time determines the maximum data rate and to ensure feature detection from the image is easier. We detected the corners using the Shi-Tomasi algorithm[88] which uses the minimum Eigen value as a scoring function. The detected corners are shown in green in the Fig. 3.2(b) where multiple points are flagged apart from the actual corners of the panel light. The four strongest points, with the highest scoring function among the detected corners are shown in Fig. 3.2(c). Since a square LED panel is used, there were four possible orders for the points depending on the orientation of the phone. The points were sorted based on the orientation from the IMU which was then compared with the known orientation of the light to detect the general direction in which the phone was pointing with respect to the light. The strongest four corners were not always the four corner points in the partial light dataset as shown in Fig. 3.1(d). In this case, we used the two strongest corners as the known corners. Depending on the axis along which the difference between the known corner points was highest, all detected corners along the axis of minor deviation between known corners were filtered and the furthest among these points from the known corners were chosen as the estimated corners labelled $P3'$ and $P4'$ in Fig. 3.1(e).

3.2.4 Positioning Technique

Once the corners are detected in the PiCS, the CCS coordinates of the same can be obtained from the pixel size and focal length. The corner points in the WCS are calculated from the known size of the light. To obtain the relationship between WCS and CCS, coordinates of the same points in both the systems are needed. The coordinates of the light corners in CCS are known to be a scalar multiple of the image corner coordinates in CCS[89] as shown below

$$P_i^C = \lambda_i \times p_i^C, \qquad i = 1, 2, 3, 4 \tag{3.1}$$

where λ_i is the scalar multiple for each corner point. To solve for these four unknowns the known sides of the light were used in [89], which works only when the entire light is

visible. When the light is partially visible, the Euclidean distance between known points of the panel light will provide one equation as follows

$$\|P_i^C - P_j^C\| = d_{i,j} \tag{3.2}$$

where P_i^C and P_j^C are the two visible corner points in the CCS and $d_{i,j}$ is the distance between corner points which will be the same here since the light is square. We also know that the sides are perpendicular, so even if the exact corner points are not known estimated corner points in the same direction as the actual corner are obtained as shown in Fig. 3.2(e). The dot product of these sides will then be zero, as follows

$$(P_i^C - P_j^C) \cdot (P_i^C - P_{k'}^C) = 0 \tag{3.3}$$

$$(P_i^C - P_j^C) \cdot (P_j^C - P_{m'}^C) = 0 \tag{3.4}$$

where $P_{k'}^C$ and $P_{m'}^C$ are the estimated corner points. We know that all four points are coplanar, which can be defined as follows

$$\{(P_i^C - P_j^C) \times (P_i^C - P_{k'}^C)\} \cdot (P_j^C - P_{m'}^C) = 0 \tag{3.5}$$

Now we have four equations in four unknowns which can be solved to find the scalar constants λ_i for the four points using which the corner points in the CCS can be obtained. The relationship between CCS and WCS is shown below

$$P_i^W = R_C^W \cdot p_i^C + \theta_C^W, \qquad i = 1, 2, 3, 4 \tag{3.6}$$

where R_C^W is the rotation matrix which provides the pose and θ_C^W is the translation vector which provides the position. We performed positioning by solving this equation using Levenberg-Marquardt optimisation technique.

3.3 Results and discussion

Some of the best camera-based single panel VLP systems are shown in table 3.2, where LED refers to the panel light shape and the full and part columns refer to fully and partially visible lights. Among the square panel techniques [85] estimated only the position and not the pose. Since [22] performed the best it was chosen as the state of the art

(SOTA) technique for comparison with the proposed technique. The proposed technique was the only technique capable of positioning with two corners and also the most accurate VLP technique for fully visible lights.

3.3.1 Orientation Identification

TABLE 3.1: Azimuthal error results

Height(cm)	120	132	144	156
Angle error(degrees)	10.85	15.63	18.86	18.74

The orientation data from the IMU on the phone is recorded along with all the images. Since the light is a square, identification of the side on the top of the image is essential for correct labelling of corner points. As there are four equal sides, if the error of azimuth detection is less than 45° the general direction in which the phone Y_P axis is pointing can be identified accurately. The error between the magnetic north from the IMU and the known magnetic north direction in degrees is shown in table 3.1, where the average absolute error for all the images at different heights from the transmitter are shown. The average error is less than half of the forty five degree threshold ensuring accurate orientation detection and hence accurate corner labelling.

TABLE 3.2: Comparison of camera-based single LED VLP systems

Ref.	***Method***	***LED***	***3D error (cm)***		***Area (m^2)***	***Height (m)***
			Full	**Part**		
[81]	circle features	○	7	-	1.8X1.8	1-1.4
[30]	circle projection	○	15.15	-	3X3	1.5-2
[83]	ellipse fitting	○	11.2	-	1.8X1.8	1-2
[82]	plane intersect	○	5.46	-	2.7X1.8	1.45-1.75
[85]	random forest	■	4	-	1.2X1.2	1.23-1.66
[84]	corner PnP	■	4.6	-	1X1	1-2
[22]	corner PnL	■	2.73	-	0.5X0.5	1.48
This work	est. corner PnL	■	0.9	2.27	1.2X1.2	1.2-1.56

3.3.2 Full LED Detection

The proposed positioning technique was applied to the full light dataset and the results were compared with the SOTA V-P4L algorithm[22]. The SOTA proposes operating the four corner LEDs of a rectangular panel light individually to beam a unique code from each of them facilitating accurate corner detection and labelling. Since this is not possible on commercial off the shelf (COTS) panel lights, they have used four different LED lights to mark the corners of a rectangular panel light. We have employed the V-P4L algorithm with the corners detected and labelled using our proposed technique. The mean 3D positioning error is the mean Euclidean distance between the actual location and the predicted location. The cumulative distribution function (CDF) of the results from the proposed technique and the SOTA are marked as such in Fig. 3.3, where the title of each individual plot is the distance between the transmitter and the receiver. The proposed technique performs better than the SOTA at all four heights and the maximum error reduces as the distance between the transmitter and receiver decreases.

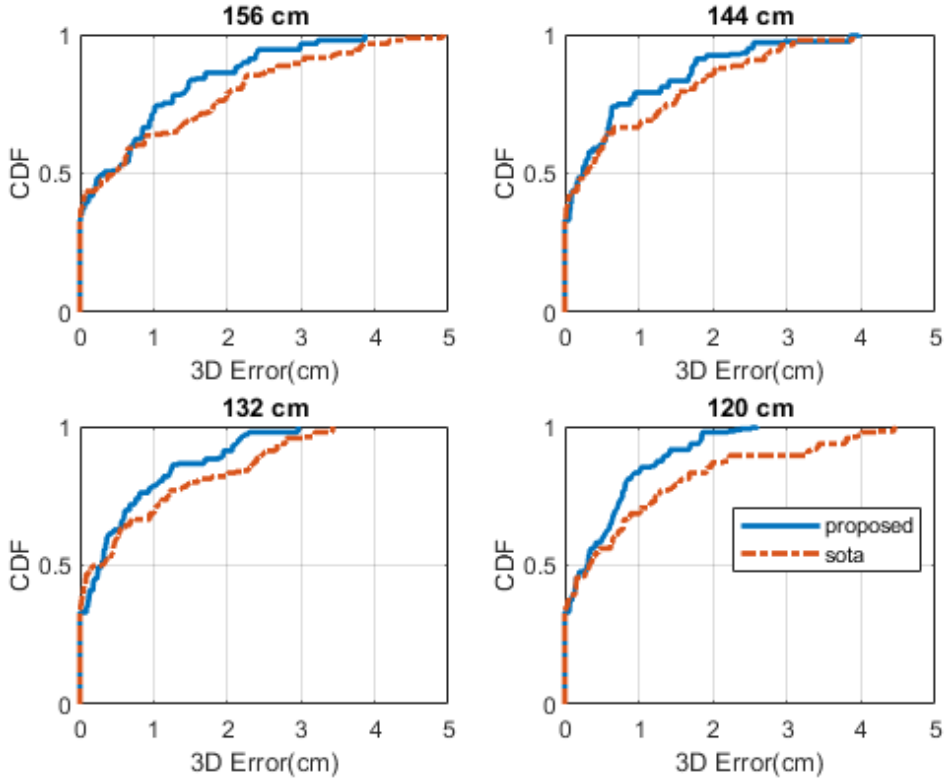


FIGURE 3.3: CDF of 3D mean positioning errors at different heights from transmitter.

The pose error is the mean absolute difference between the estimated angle and the

ground truth along the three axes. The angle error about all three axes for both the proposed technique and the SOTA is shown in Fig. 3.4, where they are plotted as a function of the distance between the transmitter and receiver. The SOTA performs marginally better or the same as the proposed technique for the x and y axes but the z axis, which is the azimuth, sees a drastic drop is accuracy from the SOTA compared to the proposed technique. There was no clear correlation between the angle errors and the height from the transmitter for either the SOTA or the proposed technique.

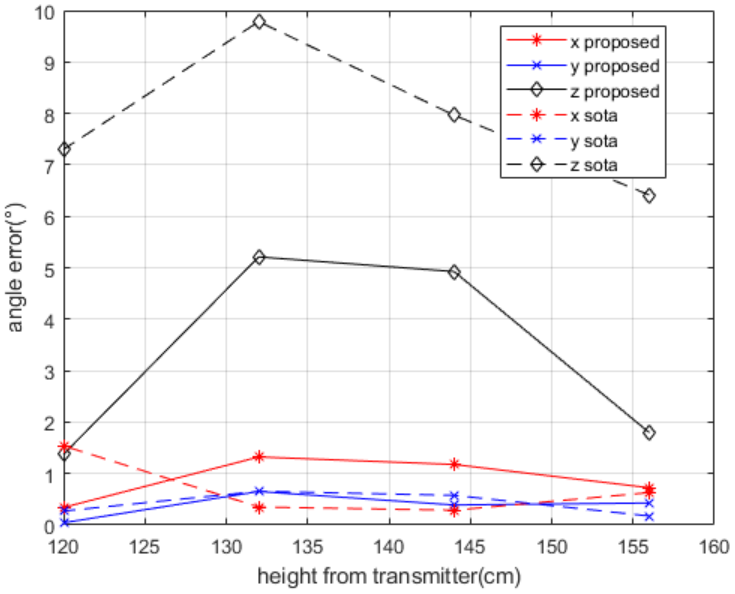


FIGURE 3.4: Angle errors of the proposed and SOTA techniques at different heights from transmitter.

3.3.3 Partially Visible LED Detection

The results of the proposed technique on the partial light dataset are shown in Fig. 3.5, where the mean 3D positioning error at each grid location is reported for the two different heights from the transmitter. The error increases with the distance between transmitter and receiver. The error at 156cm is either equal or higher than 120cm at all the twelve grid locations. Though the SOTA estimates pose and position when three corners are visible, there are no other techniques to the best of our knowledge which have performed visible light positioning when two corners are visible. Hence the results are compared with the full light dataset positioning results in the table 3.3,where the # of corners refers to the number of corners of the light visible in the image. While sub 6cm accuracy was

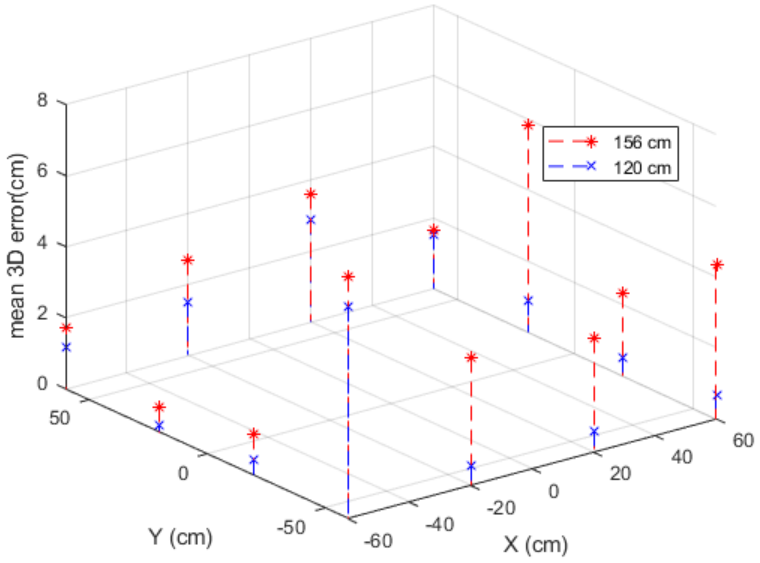


FIGURE 3.5: Positioning errors of the proposed technique on partially visible lights.

TABLE 3.3: Positioning error results

# of corners	***Height(cm)***	***3D error (cm)***	***Angle error (degrees)***		
			X-axis	**Y-axis**	**Z-axis**
4	120	0.74	0.34	0.04	1.37
2	120	1.41	2.54	3.05	4.51
4	156	1.31	0.72	0.43	1.79
2	156	3.13	2.28	4.34	2.63

achieved for the partial light dataset, the positioning error was twice as high for both the heights with the same overall pattern of increasing error with increase in height being observed. The angle error is also twice as high with no such overall pattern with the height from the transmitter. The full light dataset observed high errors only for the z axis angle while the partial light dataset performs similarly for all three axes with the errors remaining less than five degrees for the individual axes.

3.4 Conclusion

A camera based VLP technique for pose and position detection from a single partially visible LED was proposed. The light location and orientation were assumed to be known and the relative position of the receiver with respect to the transmitter was obtained using

a structure-based technique. The proposed technique was shown to perform better than the current SOTA solution on both the pose and positioning accuracy. The mean 3D error was shown to be less than 1cm when averaged across four different heights for fully visible lights. The proposed technique was shown to achieve sub six cm accuracy when only two corners of the LED were visible at two different heights from the transmitter. This can be used to address outages in indoor positioning increasing the robustness of indoor navigation solution. Though the proposed technique is better than the current state of the art technique, there still exists scope for further improvement. The heuristic sensor fusion strategy proposed in this chapter was chosen to ensure robust implementation on smartphone hardware. Deep learning models capable of handling more complex situations can be developed in the future. Environmental variation such as dynamic lighting conditions in the same building at different times and their effect on positioning can be studied in detail to improve robustness. A more general solution encompassing different light shapes, intensities and trajectories of motion along with the extent to which partially visible lights can be used for VLP can be further explored with the proposed simulation techniques.

Chapter 4

Machine learning based fine indoor positioning technique

4.1 Introduction

Visible light positioning(VLP) has gained prominence as a highly accurate indoor positioning technique. Few techniques consider the practical limitations of implementing VLP systems for indoor positioning. These limitations range from having a single LED in the field of view(FoV) of the image sensor to not having enough images for training deep learning techniques. Practical implementation of indoor positioning techniques needs to leverage the ubiquity of smartphones, which is the case with VLP using complementary metal oxide semiconductor(CMOS) sensors. Images for VLP can be gathered only after the lights in question have been installed making it a cumbersome process. These limitations are addressed in the proposed technique, which uses simulated data of a single LED to train machine learning models and test them on actual images captured from a similar experimental setup.

The indoor positioning systems (IPS) have been researched extensively both commercially and in academia owing to the wide array of applications it caters to. While there are several extant positioning techniques, VLP has a unique set of advantages, which makes it viable for further study. The indoor positioning problem consists of two steps, identifying the location of the LED and estimating the receiver location with respect to the LED. Radio fingerprinting[80] and optical camera communication(OCC)[90] have been used to solve the first part.

Several techniques have been used to estimate receiver location using VLP, of which most still use geometric processing and computer vision for localization. A single LED positioning system for circular LEDs was proposed in [30] and a similar computer vision technique was proposed in [22] for rectangular LEDs but both techniques fail when the shape of the LEDs change. While machine learning has been used for receiver tilt correction[91] and regression neural networks have been used for positioning[92] both techniques fail to provide for data augmentation and require cumbersome geometric processing for feature extraction. The use of simulation for data augmentation has been explored in [93], but they fail to take the transmitter details such as luminous intensity and receiver details such as exposure into consideration and end up with a 2D shape projection. This work proposes a single LED VLP technique using simple feature extraction to employ tree-based machine learning techniques. The dearth of data for training and cumbersome data collection was addressed through simulation, which can also be used for other deep learning models. The proposed technique was shown to outperform standard computer vision and neural network based models.

4.2 Methodology

4.2.1 Proposed Structure

The proposed structure outlined in Fig. 4.1, shows the two major parts of the technique, offline and online process. The first step of the offline process in the proposed structure is the image simulation using Blender[94], where from features are extracted. The simulation technique allows for the programmatic generation of thousands of accurate labelled images which would be time consuming to collect manually. This ensures the models for a specific building can be trained as the lighting fixtures and their locations are being planned before construction begins, which in turn can be used to optimise the lighting locations and fixtures improving positioning performance. To allow for the deployment of this proposed technique on limited hardware such as smartphones, feature extraction is used to convert the image into a list of ordered corner points, which become the input features on which the machine learning model was trained. This simple step removes the need for deep learning models which perform advanced feature extraction from unstructured data such as images. The tree-based machine learning models are then trained using the list of points as input and the 3D location as the output. In the online process,

the images of the light captured by a smartphone camera are processed to extract corners after which the trained machine learning model is then used to test performance. This produces the location of the light in the receiver coordinate system(RCS). This however

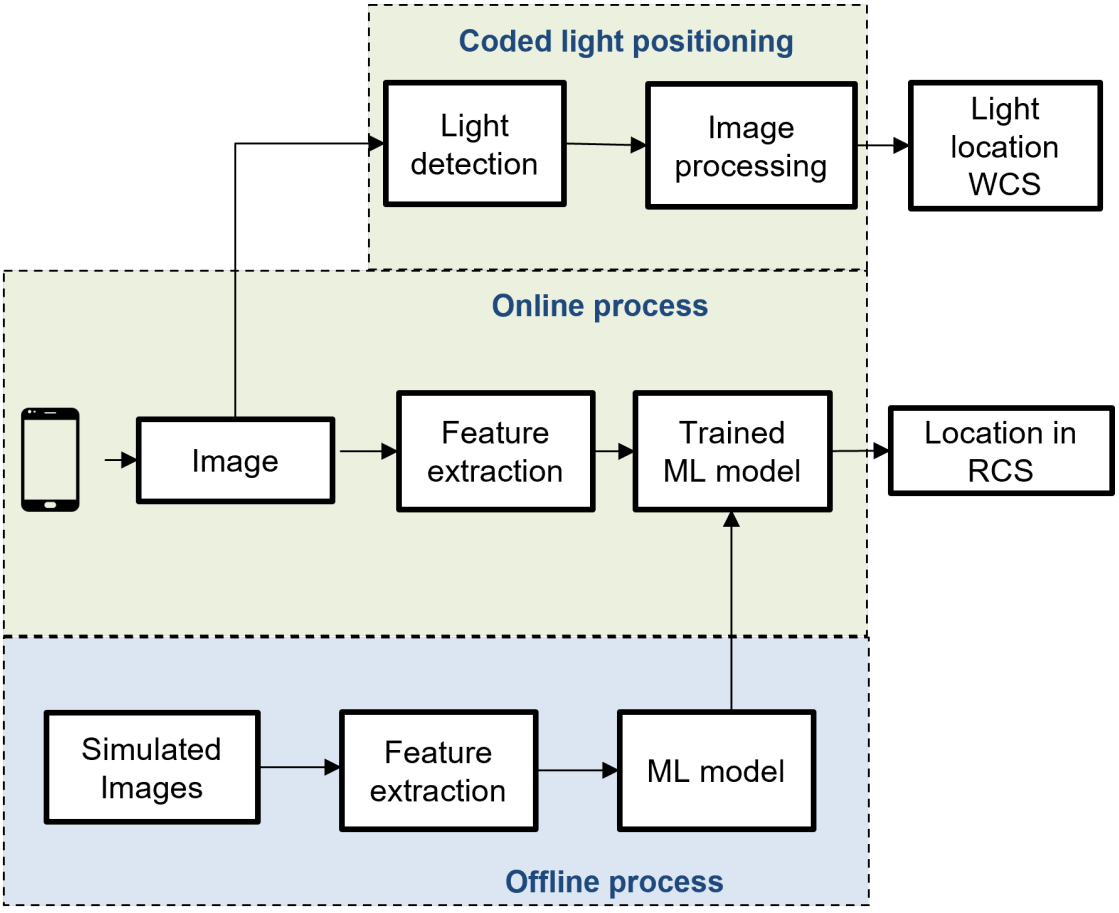


FIGURE 4.1: Overall flow of proposed structure.

is only a part of the online process, since the location of the light is needed to identify the receiver location in the world coordinate system(WCS). This is achieved using the high switching rate of LEDs. A unique ID is assigned to an LED and the ID is encoded using differential manchester encoding and beamed to the receiver using on off keying(OOK) and due to the rolling shutter effect of CMOS sensors, a temporal record of the different states of the transmitter are captured in a single image. This is then decoded using the technique proposed in [90]. The focus of this work is on 3D location estimation of the receiver with respect to the transmitter since the demodulation technique produces hundred percent detection over the range tested.

4.2.2 Experimental setup

The experimental setup for data collection to train and test the proposed technique is shown in Fig. 4.2(a), where a grid is made on the ground using tape covering 2m by 2m with each line in the grid, both horizontal and vertical being spaced 20 cm apart. This grid will act as a reference for accurate data collection using smartphones since it is done by placing the camera on the tripod with the screen facing the light. The light is 256 cm from the ground and by controlling the height of the tripod the distance from the light is controlled. The images were collected for a 1.2 m by 1.2 m grid at four different heights, 1.23 m, 1.3 m, 1.6 m and 1.66 m away from the transmitter. Here again ten images were collected at each of the 49 locations for both heights with the device orientation and tilt being changed randomly for all images to provide a wide dataset for testing generalization of trained networks.

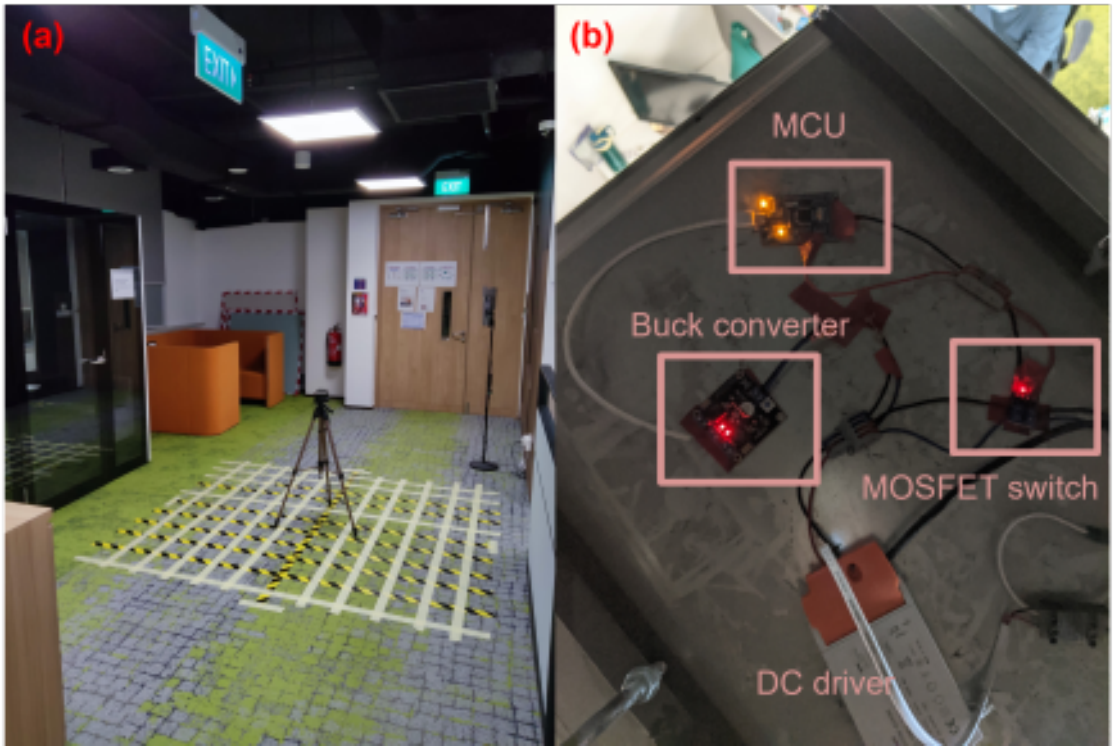


FIGURE 4.2: (a) Experimental setup (b) transmitter components

The components used to transmit the ID using the LED are shown in Fig. 4.2(b), where the STC12C5A60S2 board was used as the micro controller unit(MCU) which encodes the ID and sends the signal to the n-channel MOSFET, which turns on or off the supply from the DC driver to the LED based on the input signal. A buck converter was used to step down the LED supply to power the MCU. Since manual control of the exposure

settings was required an Android application was developed to capture images as shown in Fig. 4.3(a), where the gray scale image of the LED transmitting an ID is seen in the viewfinder. Owing to the high shutter speed, we can see the clear separation of the light and the background. The features of interest are the corners of the light, which can be extracted using the Shi-Thomasi corner extraction technique[88]. In the case of images with the transmitted ID as in Fig. 4.3(a), the image was dilated to combine the bars which can then produce corners. The parameters of interest in this case are shutter speed which determines the maximum frequency a device can decode, where it is important to note that the shutter speed must be higher than the frequency of operation since the lights are usually at the ceiling at least a couple of meters from the user and the image captured from such distances will have the light cover a small portion of the image. The other parameter of interest is the ISO, which is a measure of the sensitivity of the CMOS sensor to light hitting it. If this number is high, it will pick up low intensity lights which could lead to multipath effects from reflections due to windows or even on walls if this parameter is high enough making it difficult to identify the light bounding box in the image. These values can be modified to suit the problem space using the smartphone application developed as shown in Fig. 4.3(b). In this study using the Redmi Note 9 Pro front camera the exposure time and ISO take the values of 68 microseconds and 100 respectively throughout for all images unless mentioned otherwise. The plus and minus buttons next to the parameter on the camera settings overlay to be changed can be pressed to change them and the current value is displayed between the buttons.

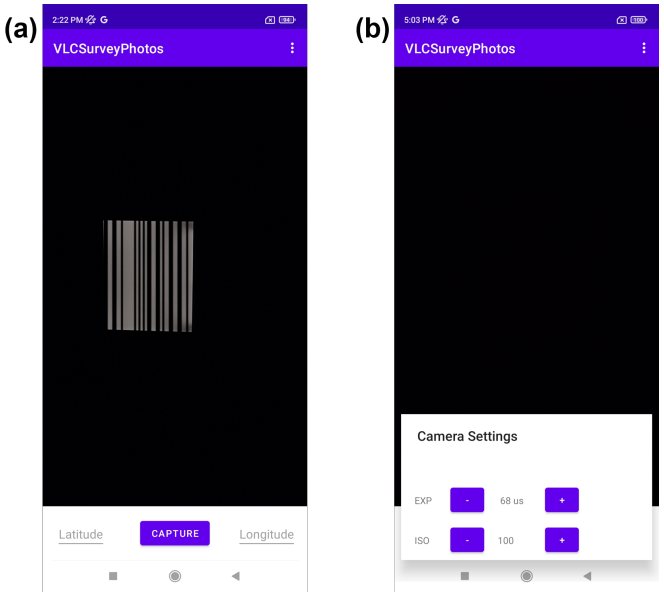


FIGURE 4.3: Receiver Android application

The next step is capturing the images which can be done by clicking the capture button on the bottom of the screen. A sample of an image in the view finder of the application is shown in Fig. 4.3(a), with the parameters set at the aforementioned exposure time and ISO values which yields a clear image of the strides of an eight-bit long code at ten kilo hertz frequency. The captured image can be named by entering the location coordinates on both sides of the capture button since this also serves the purpose of image collection for training and testing data in the case of intra-cell positioning using a single transmitter. The captured image is then to be processed to decode the location ID being transmitted by the LED, which can then be matched to a database of known location IDs to identify the transmitter location. For each of the grid locations ten images were captured at each height of which two were selected as test data and the remaining eight were used as training data.

4.2.3 Image simulation

Data augmentation techniques are generally employed in standard deep learning-based classification problems. These range from scaling, rotating to inverting images which in this case would make the image unusable. However, this is one of the challenging applications for data generation since it will be a three-dimensional regression problem eventually when defined as a camera relocalisation problem with a six degree of freedom quaternion as its output. Since collection and labelling of images accurately is time consuming and a seemingly endless amount of data can be collected depending on the accuracy of detection expected simulation using Blender was used to augment data collection.The simulation screen from Blender is shown in Fig. 4.4, where an area light was modelled to replicate the specifications of the LED used for testing. A 59.5 cm square LED panel from Lite Unite, DWUGR606036 was used for testing producing 3600 lumen with a color temperature of 4000K. The area light was modelled as a plane with an emission shader as shown in the bottom panel, where the color temperature was replicated using a blackbody node with the temperature set to the appropriate value and the polar curve of luminous intensity was used to produce an illuminating engineering society(IES) file to model the throw pattern. The IES node was used to set the appropriate signal strength of the emission shader.

The image simulated from this technique is shown in the left panel in Fig. 4.4, where the black background is obtained by setting a low exposure value. Since the shutter speed for

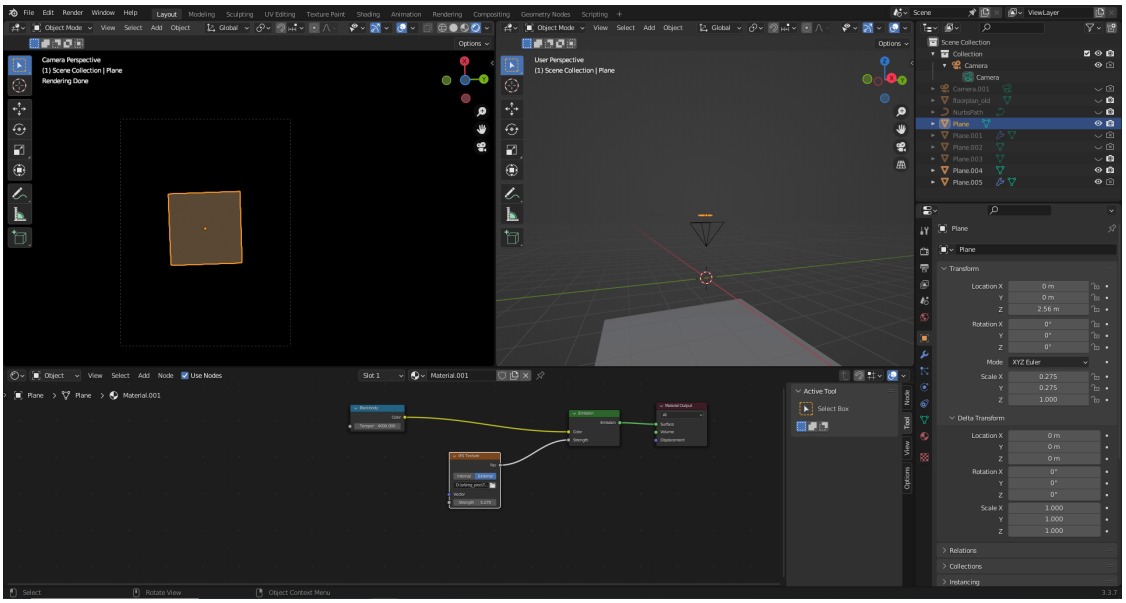


FIGURE 4.4: Blender simulation screen

VLC and reading the ID from the coded light module must be very high only the brightest parts of the image, which in this case is the light, are seen with all background features being lost. This also ensures that feature extraction from the image becomes much easier owing to the simpler image and also enables reuse of the image for all lights with the same shape since the background features are ignored. However, the pattern formed from images transmitting the ID is not simulated since the corners of the lights are the only features being used for training the machine learning model. The camera used for rendering images in Blender was placed at different positions and at different orientations in the space below the light controlled by location and rotation values. The images were generated with a 3:4 aspect ratio, which is the most common choice for smartphone sensors, at the same resolution, 1728 X 2304, as the test image to ensure compatibility between simulated and test data. The data was simulated at the same grid locations at two of the same heights 1.3 m and 1.66 m from the light with two more heights different from the experimental data capture at 1.76 m and 1.56 m from the transmitter, which will be used to test the generalizability of the trained model. The receiver orientation was swept in complete circles on roll, pitch and yaw values in increments of 45 degrees with only the images where all four corners of the light were in the FoV of the camera were retained, which produced 7982 images for all four heights. Since these images were simulated, the corners were also labelled using ray casting to be the appropriate points corresponding to the LED corners which can be cumbersome in the actual data gathering process. The LED panel used here is a square and without any background features the images will

look similar along any of the four sides leading to erroneous results without additional information. The simulation process simplifies this allowing the 3D position to be estimated without pose or orientation information. While the technique proposed in this chapter only uses the corners of the light as input features for the machine learning model and those can be projected from equations as discussed in [30], the simulation technique allows for future expansion when advanced features are needed in the case of different lighting fixture shapes or multiple lights in the field of view of the camera. The ability to simulate images from a 3D model allows for generation of images particular to a specific building even before construction begins. With the Singapore government making building information models (BIM) mandatory for new projects[95], detailed building models with specific information about the materials in all parts ranging from windows to walls, the accurate IES files of lighting fixtures can allow for programmatic generation of labelled images for training models from scratch or for fine-tuning pre-trained models with the easy to use Blender add-on. Blender being an open source free to use software provides additional justification for the proposed technique.

4.3 Results and Discussion

4.3.1 Model selection

The corners were extracted from both the simulated and real datasets and ordered with the list of points ordered in the same sequence manually in the case of real images with simulated images being generated with the corners labelled. The real dataset will be used as the test set in this section. Here, the images at 1.3 m and 1.66 m from the transmitter were split into train and test sets with 2 images at each grid location for the latter and the rest for the former. The test set has 196 images and train set has 784 images in the real image dataset. All the simulated images were used for training, which was 1163 at 1.3 m and 2516 at 1.66 m, totals to 3679 images. The simulated dataset is 4.6 times the real dataset owing to the ease of collecting and labelling data. Three models were trained on the real dataset and tested at 1.66 m from the transmitter. Two tree-based ensemble techniques, random forest[96] and extreme gradient boosting(xgboost)[97] were tested along with a multi layer perceptron[98].

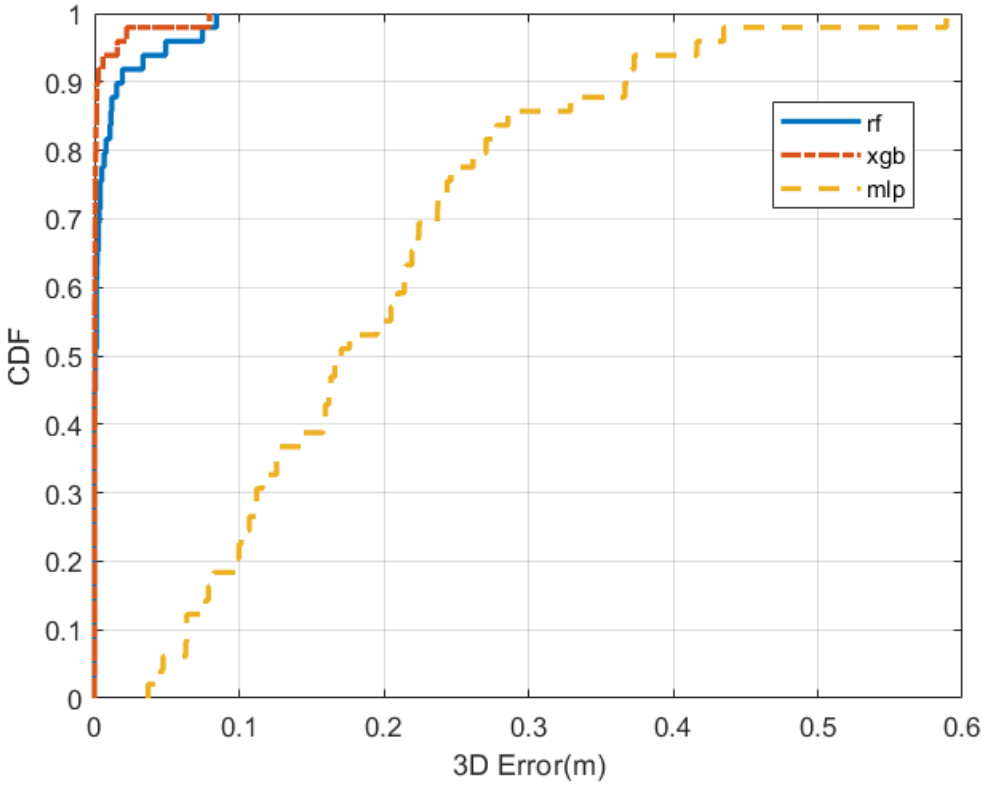


FIGURE 4.5: CDF of 3D positioning error for different models

The 3D positioning error is the Euclidean distance between the estimated location and the actual ground truth. The cumulative distribution function(CDF) of the 3D positioning errors is shown in Fig. 4.5, where the tree-based techniques outperform the neural network. The models were implemented using the scikit-learn package[99], with default values for all parameters apart from number of estimators for the tree-based techniques, which was changed to 150 and for the neural network five hidden layers with 200 nodes in each were used. The 3D positioning error using the neural network for 90% of points is shown to be less than 40 cms while random forest, which is the worse of the tree-based techniques, has all points less than 10 cms. The marked improvement is expected since the images are converted to a list of points making it a structured dataset. The tree-based techniques are shown to outperform neural networks across multiple structured datasets[100].

Among these tree-based techniques, the mean 3D positioning error at each grid location is shown in Fig. 4.6, where the xgboost model is marked with asterisk and the random forest model is marked with cross. The results of the random forest model shows that most of the error comes from outermost points in the grid and some from points closer to the center, with multiple points producing more than 5 cm of error. In the xgboost model

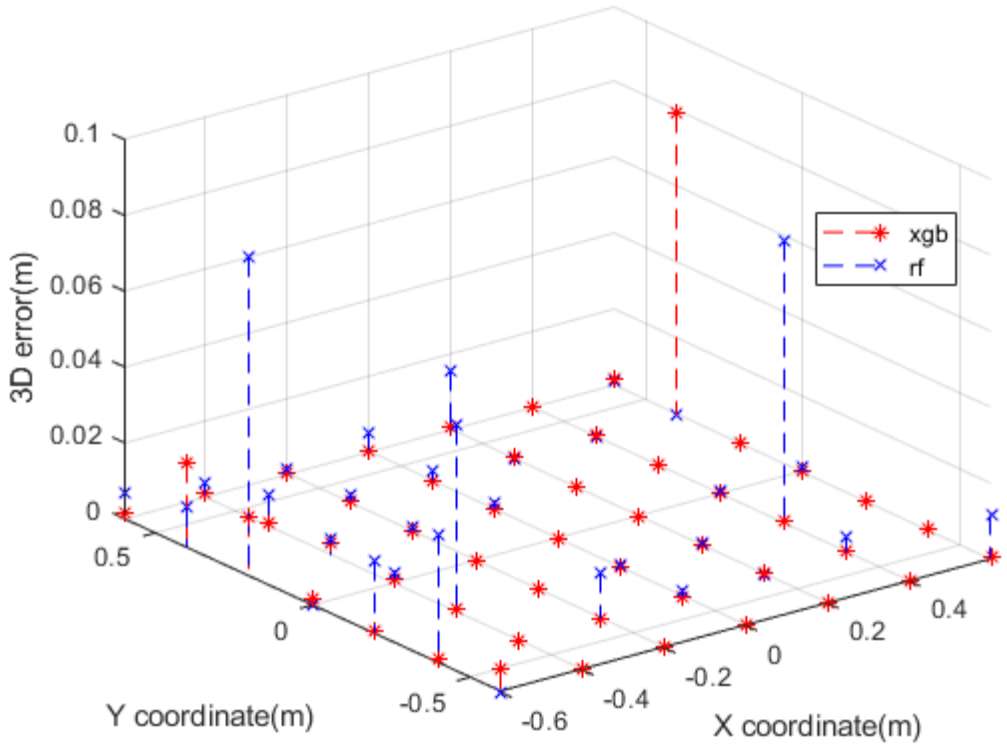


FIGURE 4.6: 3D positioning error for tree-based models

all the errors are in the outermost points with only one point producing more than 5 cm error. This explains the better overall performance in the case of the xgboost model over the random forest model. The 3D error CDF also shows that though both models have similar maximum errors, the error for xgboost is lower across all the points in the test dataset. Thus, the xgboost model was chosen for subsequent testing.

4.3.2 Effect of simulated data

The simulated dataset was used to train a xgboost model, which was tested on the real dataset. The results of the same are to be compared with a xgboost model trained and tested on real images and the closest competitors technique used on the real test dataset. The closest competitor is marked sota in Fig. 4.7 to indicate the state of the art(sota) results reported by the same[22]. The sota uses computer vision to identify geometric relations between the four points in the image plane, camera coordinate system and world coordinate system. They also use a photo detector(PD) to identify the location of the light in the world coordinate system.

TABLE 4.1: Mean 3D positioning error results

Train	***Test***	***Distance from light (m)***	***Mean 3D error (cm)***
Simulated	Real	1.66	3.11
		1.3	2.65
		1.6	5.32
		1.23	4.9
Real	Real	1.66	0.29
		1.3	0.104
		1.6	1.32
		1.23	1.17
state of the art		1.66	6.35
		1.3	6.17
		1.6	8.13
		1.23	6.07

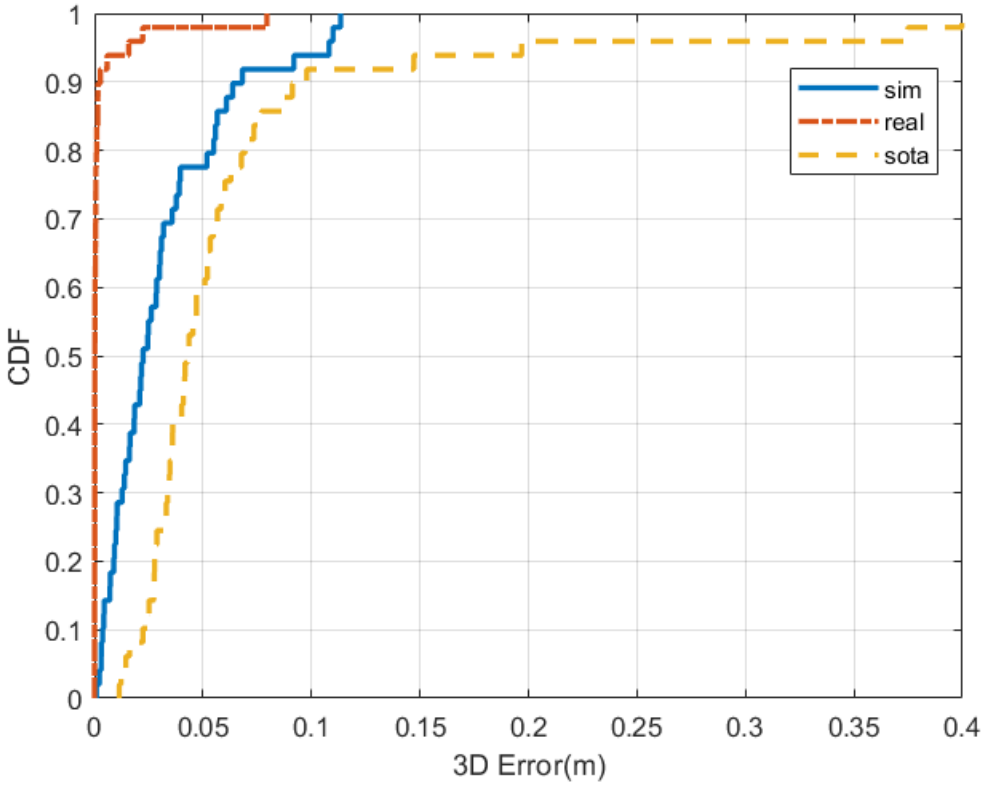


FIGURE 4.7: CDF of 3D positioning error for data at 1.66 m

The model trained on the real dataset produces the best results of the three techniques though it was trained on four times fewer data points as observed from the CDF of 3D error at 1.66 m from the light in Fig. 4.7. This however, fails to take into consideration

the time intensive labelling process of the corner points. The maximum error produced here is less than 10 cm when trained and tested with real images. The maximum error rises to 12 cm in the case of training with simulated images and testing with real images. The sota performs the worst with maximum errors of upto 40 cm. However more than 90% of the points have less than 10 cm of error in all three cases.

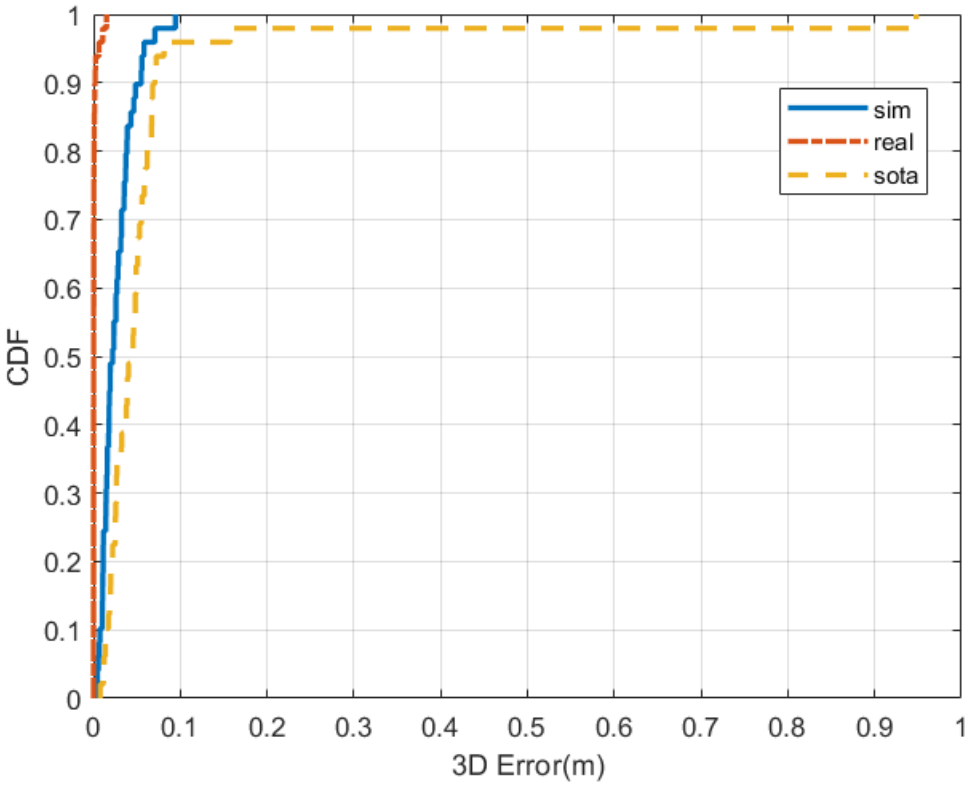


FIGURE 4.8: CDF of 3D positioning error for data at 1.3 m

The CDF of 3D positioning error at 1.3 m from the light is shown in Fig. 4.8, where the maximum errors for both models trained on simulated and real images decreases but the maximum for the sota increases to 90 cm while the 90% performance improves slightly indicating that there are outlier grid points in the sota affecting the overall performance. The mean 3D positioning errors are listed in table 4.1, with the train and test columns indicating the training dataset and testing datasets used. The mean errors are consistent with the CDF observed at these two heights, as we move closer to the light the overall positioning error decreases. Though the positioning accuracy achieved in the simulated dataset is lower than the real images, it still is better than the sota by more than 3 cm for both the heights indicating the similarity of the simulated images to real images.

4.3.3 Performance generalization

The results reported thus far have used either simulated or real images from the same heights for training the models. However, this is not a good indicator of the model having learnt the relationship between corner points of the light in the image and the 3D coordinates of the receiver location with respect to the light. In order to test if the model has learnt this relationship two different heights of the real images were used as test datasets at 1.6 m and 1.23 m from the light. These are just 6 cm and 7 cm away from the original training locations, in order to truly test the generalization of performance on the height axis, two more sets of images were simulated at 1.76 m and 1.56 m from the transmitter. Though the test set at 1.6 m is still close to one of the datasets, the 1.23 m test set can be used to gauge consistency of results since it is further from both the datasets. The number of images for this simulated set has changed owing to the change in distance from the light, with 2683 images at 1.76 m and 1620 images at 1.56 m, to 4303 images. The real dataset however was kept at 1.66 m and 1.3 m owing to the difficulty in data collection and labelling. Since the proposed technique involves the use of simulated images rather than real images a new real image training set was not created for the new heights at which images were simulated.

The CDF of the 3D positioning error for test data at 1.6 m from the light is shown in Fig. 4.9, where the model trained on real images performs the best with the maximum error still lower than 10 cm. The model trained on simulated data shows a marked decrease in performance owing to the new dataset further away from test points and has a higher maximum error than the sota in this case. However, 90% of the points have less than 15 cm error in the simulated dataset while the sota has the same mark at less than 25 cm. This marked difference in performance is observed owing to the distance from the light being higher for this dataset and the simulated points being further away on average from the test points.

The CDF of 3D positioning error in the case of test data at 1.23 m from the transmitter is shown in Fig. 4.10, where the model trained on the simulated dataset performs better with a marked reduction in the maximum error from 25 cm to less than 18 cm. The sota achieves similar maximum error but more than 90% of the points are observed to have an error of less than 10 cm which once again indicates outliers in the case of sota causing performance issues compared to the simulated results in [22].

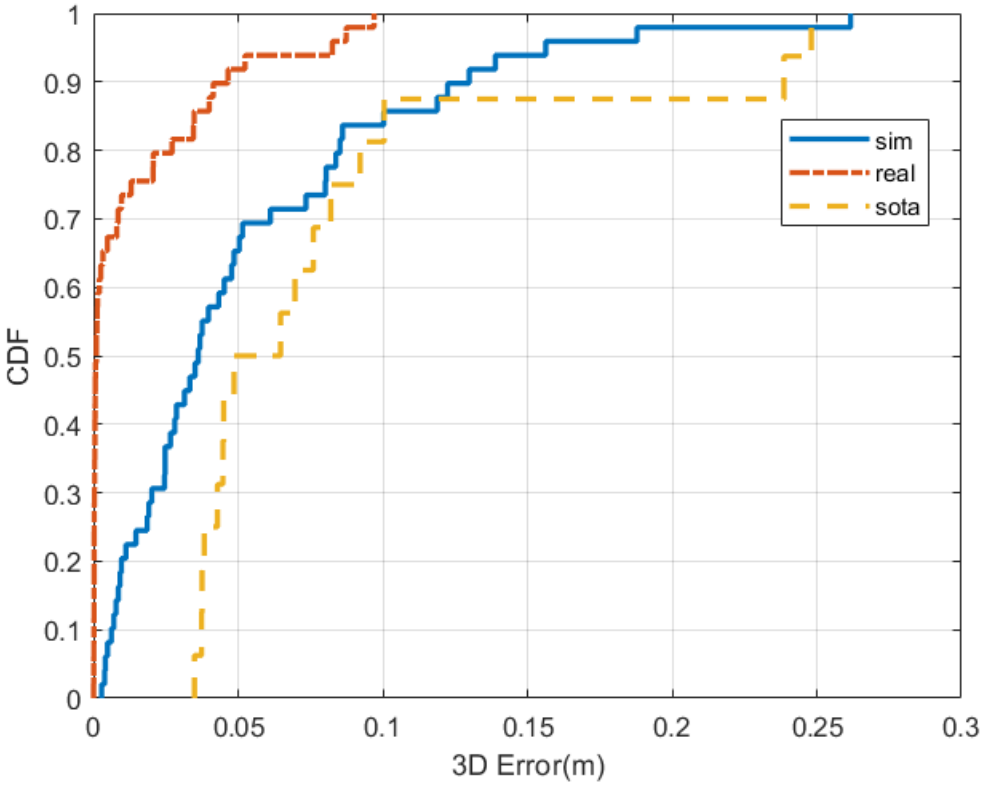


FIGURE 4.9: CDF of 3D positioning error for data at 1.6 m

The CDF of individual deviations of the estimated locations from the ground truth on all three axes is shown in Fig. 4.11, where the sota performed worse on all three axes. The simulated data observes a higher error on the x and y axis than the z axis indicating the robustness of the relationship learnt by the model. The x axis produces highest error for all three techniques with only the model trained on real images managing a 90% mark less than 5 cm error. In both the other axes almost no error is observed in the real model, but the simulated model performs better than the sota in all three axes individually producing the lowest error in the z axis.

The CDF of individual deviations for the data 1.23 m from the light is shown in Fig. 4.12, where the z axis error is the lowest for all three models owing to the receiver's proximity to the light. The maximum errors are produced in the x axis once again but this time both sota and the simulated model perform much closer to the real images model on the x and z axis with the y axis producing the highest difference between them. From the table 4.1, the mean positioning error is also consistent with the observed results thus far, the model trained on real images performs the best across the board for all heights but this data collection strategy is not scalable when applying to deep learning models. The

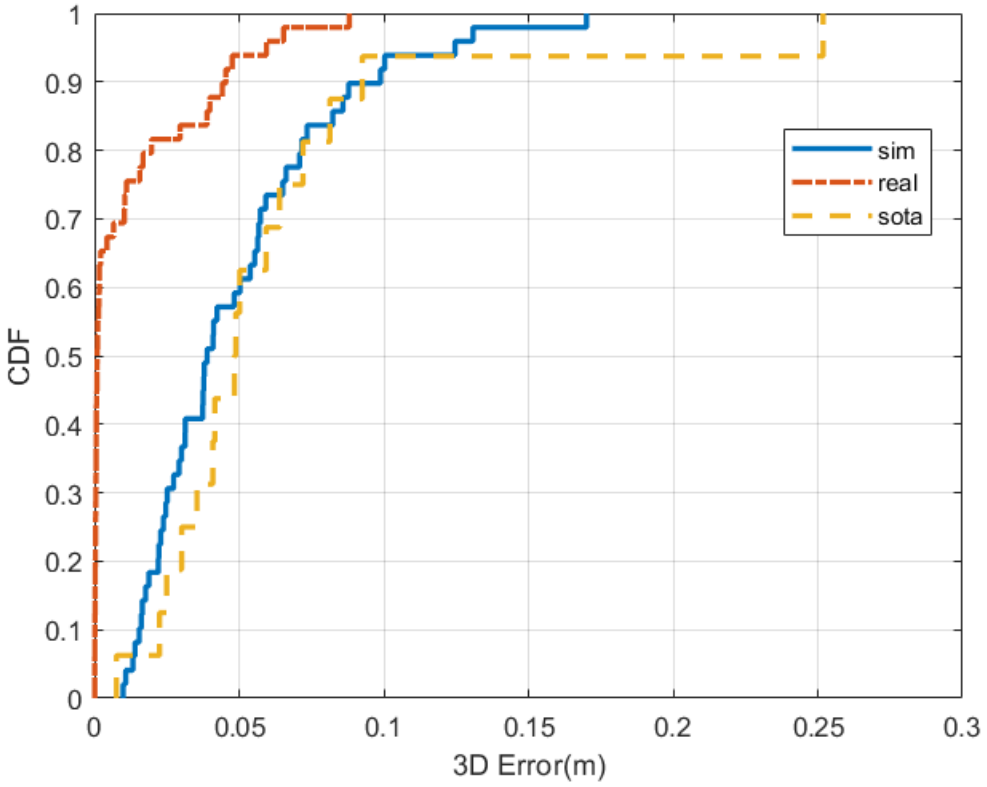


FIGURE 4.10: CDF of 3D positioning error for data at 1.23 m

simulated models perform much better than the sota by 3 cm at 1.6 m and 2 cm at 1.23 cm, which apart from the proximity to the transmitter is also driven by anomalies in the test dataset at 1.6 m which causes an increase in error across all three models but the most pronounced errors in sota. This cannot be due to height generalization testing since the sota employs a computer vision based technique and does not rely on data for modelling.

4.4 Conclusion

We proposed a tree-based VLP technique using simulated data for single LED indoor positioning without the need for data collection and labelling. The model trained on simulated images was shown to perform better than the closest competitor and within 3 cm of mean 3D positioning error from the model trained on real images. The similarity of results obtained between the simulated and real images indicates the photorealism observed in the simulation. The conversion of images to a list of points reduces the unstructured images to structured data enabling the superior performance compared to

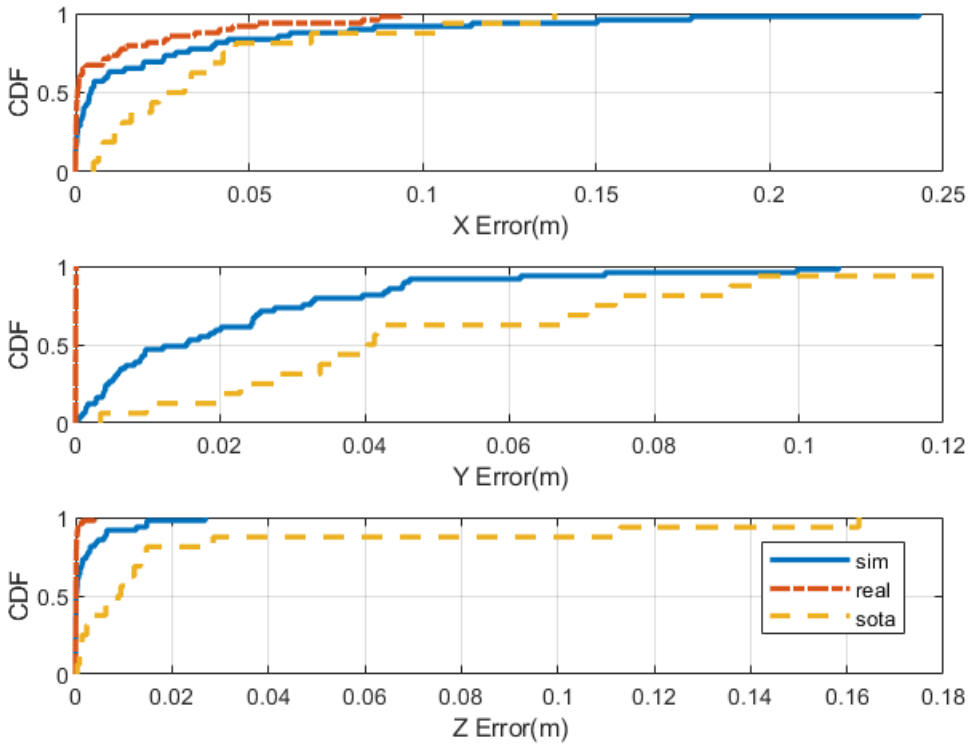


FIGURE 4.11: CDF of three axis errors for data at 1.6 m

the closest competitor. The superior performance of tree-based models on structured data is leveraged to obtain these results. The generalization of the models was tested by simulating images further from the test points and the models were shown to perform best on the z-axis with the lowest error among the three axes. The proposed technique was implemented with a focus on deployment in a smartphone, which severely limits the complexity of a model that can be trained. The feature extracted for use in the proposed technique simplifies the simulation requirement. However, for more complex models which use more advanced features extracted from the entire simulated image, further analysis of similarity between the simulated and real world images can be performed in the future. The dynamic environmental changes such as lighting conditions within the test environment and their influence on the images can be analysed in detail. The simulated images can be used to train more advanced, deeper neural networks to capable of handling multiple LEDs in the camera field of view. This would allow for complete utilisation of the proposed simulation technique considering the amount of data required for such models.

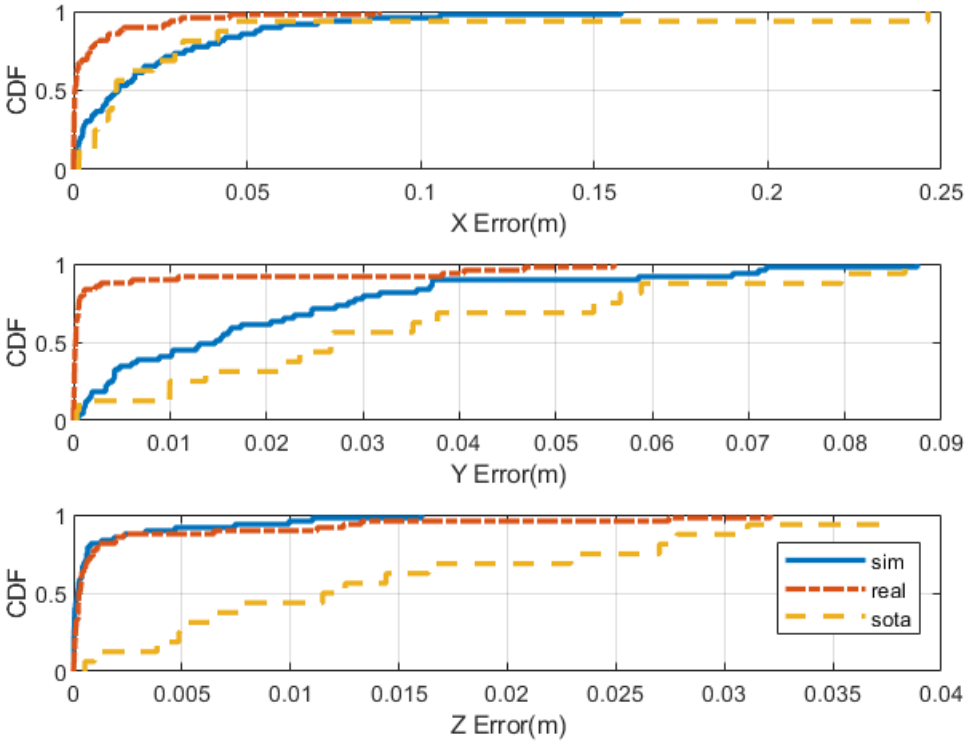


FIGURE 4.12: CDF of three axis errors for data at 1.23 m

Chapter 5

Optical Camera Communication for coarse indoor positioning

5.1 Introduction

The existing standards for VLC systems ranging from IEEE 802.15.7 to 802.11bb prioritize the use of PDs over cameras, owing to high speeds achieved by the former. However, the ubiquity of cameras provides important use cases for camera based VLC such as indoor positioning[101, 102] and vehicular communication[103, 104]. Cheap CMOS sensors can be used for unidirectional communication to facilitate indoor positioning. Coarse transmitter localisation can be achieved using radio fingerprinting[79] or magnetic field strength based techniques[27], but they are not as accurate as camera based VLC. To identify each transmitter uniquely, the length of code being transmitted has to be high, which requires a high data rate. Simulation tools are generally used to identify the maximum frequency of operation and hence decide the maximum number of bits in the code. While PD based VLC has several tools to identify this, camera based VLC does not.

Current OCC simulation techniques are not as accurate as in the case of PDs, owing to the different processing pipelines employed by CMOS cameras. A radiometric approach with a complete image processing pipeline was produced to test camera performance[105] but the rolling shutter effect of CMOS cameras was not included. A Lambertian model for the transmitter was used to incorporate distance into the simulation[106] and transmitter illumination data was used in Blender to generate photorealistic images of lights[85] but

both lack OCC capabilities. CamComSim[107] employs a Markov-modulated Bernoulli process to simulate a network and produce probability of success but it does not generate an image to facilitate decoding algorithm testing. OCC simulation has been performed using DC gain of each pixel in the area of view of the camera[108] and using photometric properties[109] but both do not outline operation beyond shutter speed with the former adhering to the Nyquist rate. Though there are simulation techniques available for OCC, they do not work for frequencies beyond the shutter speed.

Most camera based VLC simulation techniques use physical principles such as radiation or photometry to determine equations for performance metrics such as signal to interference plus noise ratio or maximum bit rate directly which leads to the rigidity of these techniques. We seek to address this issue using a simple weighted average of expected light intensities over individual exposure periods allowing for the simulation of images close to reality. To test the accuracy of the proposed simulation technique, simulated images were decoded using a commonly used thresholding technique[89, 110] and the results were compared with experimental data.

The main contributions of this work are

- Proposed a simple technique to simulate OCC at any signal frequency irrespective of exposure time.
- Compared discrete Fréchet distance of proposed simulation technique and SOTA technique.
- Experimental validation of simulation for two different cameras and two different transmitters.
- Tested influence of noise on a conventional detection technique for different transmitter and receiver properties.
- Proposed an improved demodulation technique by training machine learning models on simulated images.
- Analysed BER to test influence of modulation and encoding schemes on the proposed demodulation technique.

5.2 Methodology

The process outline is delineated in Fig. 5.1, where the transmitter is an LED panel light. A 10 bit code was chosen since it will allow for 1024 unique variations which can be assigned to as many lights in the case of transmitter localisation for indoor positioning. This code is then encoded using differential Manchester encoding to limit the run length of same bits since this will cause noticeable flicker at lower frequencies. We have used on off keying (OOK) modulation, which is the most commonly used modulation technique for OCC[108]. Since image processing is computationally intensive, this simple modulation scheme allows for detection on lower end smartphones. By capturing images of the LED panel through a camera the transmitted data is received. To simulate this received image, the camera parameters such as the exposure time, focal length and aperture along with the transmitter properties such as the area of the panel and luminous intensity are used. Samples for the simulated and actual image for the same parameters are shown in Fig. 5.1, which look similar. While we can compare the images directly for similarity metrics, it does not tell us if the simulated image will perform the same as the actual image on detection algorithms which is the main use of simulation. Hence, we perform image processing to get the transmitted string which is then decoded to obtain the transmitted code. We compare the received codes with the transmitted code to find the number of correct bits which is the success rate of transmission.

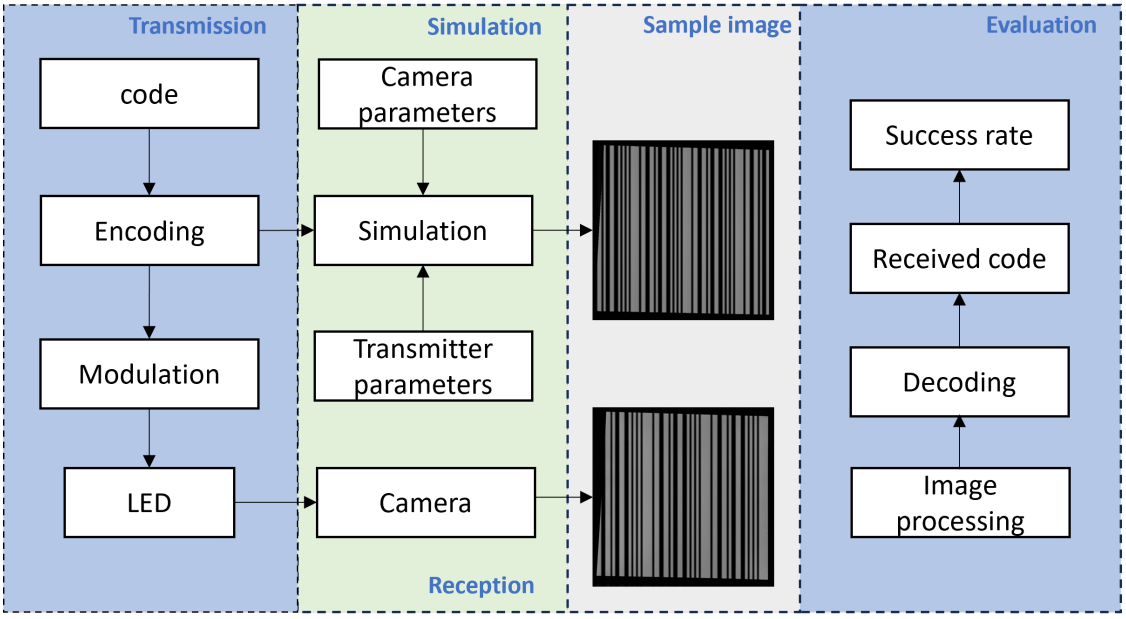


FIGURE 5.1: Process outline.

5.2.1 Experimental Setup

To test for the robustness of the proposed simulation scheme, we collected images on a phone and tablet front camera at different distances from the transmitter and at different switching frequencies. We use the front camera on both these devices since it allows the user to look at the screen of these devices for navigation when using the lights above them for positioning. The transmitter and receiver specifications are mentioned in table 5.1, where the models of the phone and tablet used are also outlined. The code to be transmitted was encoded on an Arduino Uno and an n channel MOSFET switch was used to switch the light on and off according to the encoded bit string at switching frequency within the range outlined in table 5.1. To facilitate ease of data capture we placed the light on the floor and the receiver at distances ranging from 60 cm to 200 cm in 20 cm increments. For each frequency, exposure time and distance, five images were captured.

TABLE 5.1: Device specifications

Parameter	**Specification**
Transmitter model	Lite Unite LED panel, DWUGR606036
LED shape and size	59.5 cm square panel
Colour temperature	4000 K
Luminous flux	3600 lm
Transmitter model	LD-2835-R-12W
LED shape and size	18 cm round panel
Colour temperature	6500 K
Luminous flux	880 lm
Code	1110111000
Frequency	2 to 20 kHz
Modulation	OOK
Encoding	Differential Manchester
Phone model	Redmi Note 9 Pro
Image resolution	2304×1728
Exposure time	68 μs & 136 μs
Readout time	8 μs
Aperture	2.25
Tablet model	Galaxy Tab S7
Image resolution	2448×2448
Exposure time	52.5 μs & 105 μs
Readout time	13 μs
Aperture	2

5.2.2 Simulation Technique

The image generation pipeline for CMOS sensors is the same across the current simulation techniques for OCC. The amount incident photons on the sensor is determined by the luminous intensity of the transmitter along with the lens aperture and exposure time of the camera. These photons are converted to electrons based on the quantum efficiency of the sensor and the resulting voltage is amplified based on the ISO speed of the camera. This is then digitized and converted to a pixel value value between 0 and 255 through gamma encoding. Each camera processes these pixels through a unique image processing pipeline which is not revealed to the user, following which the final image is generated. The general pixel value determination for when the light is on or off is performed as outlined in [109], and detailed below as PV_{max} and PV_{min} respectively.

$$PV_{max} = 118 \left(\frac{S \times t}{K \times N^2} \left(L_v + E_v \frac{R}{\pi} \right) \right)^{1/\gamma} \tag{5.1}$$

where S is the ISO speed of the camera set to be 100 in our experiments, t is the exposure time, N is the lens aperture, L_v is the luminous intensity of the transmitter which was 3600 cd/m^2 in our experiments, E_v is the external illuminance measured to be 290 lux, R is the reflectance assumed to be 40% as per [111], K and γ are constants assumed to be 12.5 and 2.22 as per [109].

$$PV_{min} = 118 \left(\frac{S \times t}{K \times N^2} \times E_v \frac{R}{\pi} \right)^{1/\gamma} \tag{5.2}$$

In a CMOS sensor with rolling shutter, each column is exposed individually for the exposure time. Each column aggregates the amount of light over this time period. When the switching period is less than the exposure time, the light will be on and off within a single exposure period. We calculated the simulated pixel value for each column PV_{sim} as the weighted average of PV_{max} and PV_{min} using the duration for which the light is on and off as the weights. The formula for which is as follows

$$PV_{sim}(i) = \frac{PV_{max} \times t_{on}(i) + PV_{min} \times t_{off}(i)}{t} \tag{5.3}$$

where $t_{on}(i)$ is the duration for which the light was on when column i was exposed and $t_{off}(i)$ is the corresponding duration when the light was off. There is a small delay between when each column is exposed called the readout time, which is much smaller than the lowest possible exposure time for the camera. Therefore there is a significant overlap

in the states for consecutive columns. This allows us to observe the light states even when the exposure time is greater than the switching period. The resolution and the readout times of the cameras used are listed in table 5.1. Since we do not know the exact image processing steps used to arrive at the final image, we employed a commonly used contrast enhancement technique called histogram equalisation to improve detection.

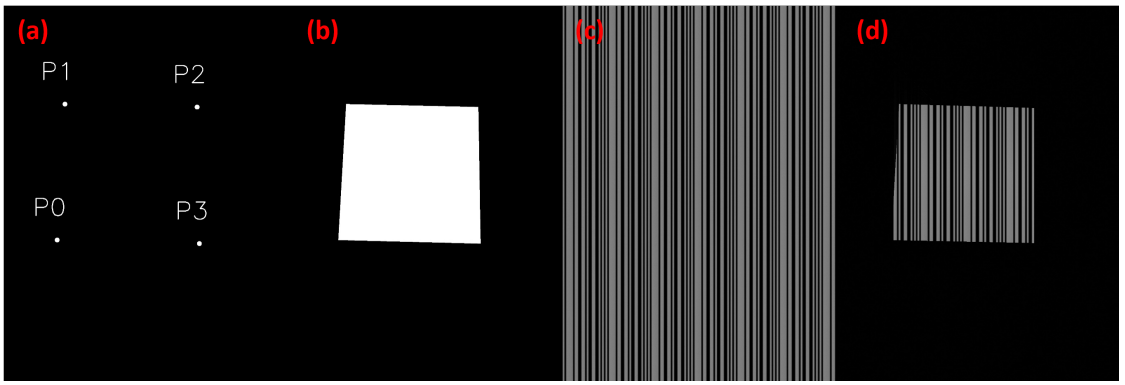


FIGURE 5.2: (a) projected light corners (b) mask of the light area (c) calculated pixel value (d) simulated image

The pixel value calculation uses transmitter and receiver properties but does not take the distance between them into consideration. The larger the area covered by the transmitter in the image better the chances of detection. Using the intrinsic camera properties we projected known coordinates of the LED corners to the image plane, which are labelled in Fig.5.2(a). These corners are then joined to form a mask of the image area covered by the transmitter shown in Fig. 5.2(b). The string of pixel values after histogram equalisation is shown in Fig. 5.2(c), where the light covers the entire image area. By masking the area of interest we arrive at the final simulated image shown in Fig. 5.2(d).

5.2.3 Decoding Technique

5.2.3.1 Conventional Technique

To test the accuracy of simulation, we propose to compare the detection success rate of simulated and experimental values. The conventional demodulation technique outlined is based on Otsu's thresholding which was shown to be useful in similar decoding problems[110]. We identify the area of the image covered by the transmitter through image processing. We determine the brightest pixels in the image covering five percent of the area, which gives us some of the pixels in the transmitter image. If there are multiple

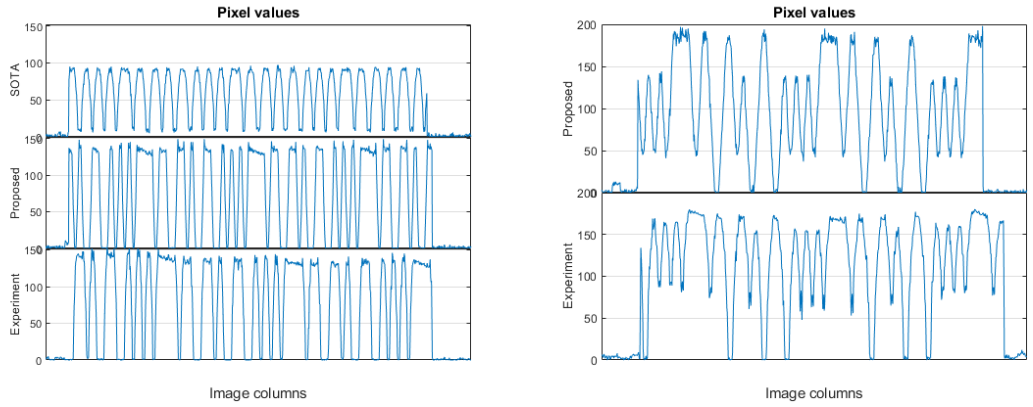
contours the overlapping area is used to grow the contour until only one contour remains. A bounding rectangle is constructed over this contour which is sliced from the original image as shown in Fig. 5.1. The columns within this area are averaged to obtain a signal. Otsu's thresholding is used to binarise the image. A histogram of the run length ones and zeros is constructed from which once again Otsu's thresholding is used to determine the average run length for ones and zeros. We used a header with three continuous ones, since the code can never have that owing to differential Manchester encoding. The first two instances of the header are used to separate the code of interest, which is then decoded to get the received code. Since the data rates achieved by OCC are much lower than VLC and the difference between a single transmitted and received code is the metric of importance, we report the success rate as a percentage instead of the bit error rate, which is defined as follows.

$$SR = \frac{\#B_c}{\#I \times \#B_t} \times 100 \tag{5.4}$$

where $\#B_c$ is number of correct bits in the received code, $\#I$ is the number of images and $\#B_t$ is the total number of bits in the code.

5.2.3.2 Proposed Technique

Neural networks have been shown to be capable of performing demodulation [112, 113]. We propose to use the simulated images to train a machine learning classifier that uses the output of the previous symbol as an input to the current symbol. Using images directly will require computationally intensive deep learning models, which we have avoided using image processing to convert the images to a string of values. We identified the bounding rectangle encompassing the light in the image using the conventional technique. We averaged the values over this bounding rectangle to obtain a string of pixel values. The header was identified within this string and the bits between consecutive headers were split into individual bits. Since the problem has been reduced to a classification problem with two output classes, simple machine learning techniques can be used to perform demodulation.



(a) less than the switching period (b) greater than the switching period

FIGURE 5.3: Pixel values when the exposure time is

5.3 Simulated Images

To show that the proposed simulation technique is accurate and better than extant techniques, we compared it to a state of the art (SOTA) simulation technique. The photometry based simulation technique in [109] is considered SOTA since it defines simulation until the exposure time is less than the switching period while others define simulation when the exposure time is less than half of it. The SOTA defines the complete band and transition band lengths to determine the sequence of pixel values for all columns.

$$h_c = \frac{t_{LED} - t}{t_r} \qquad h_t = \frac{t}{t_r} \tag{5.5}$$

where h_c is the number of columns that will be at the zero or one state, h_t is the number of columns when moving from one state to another, t_{LED} is the switching period of the LED and t_r is the readout time of the camera. When the exposure time is equal to the switching period the complete band becomes zero as per this technique.

5.3.1 Pixel Value Comparison

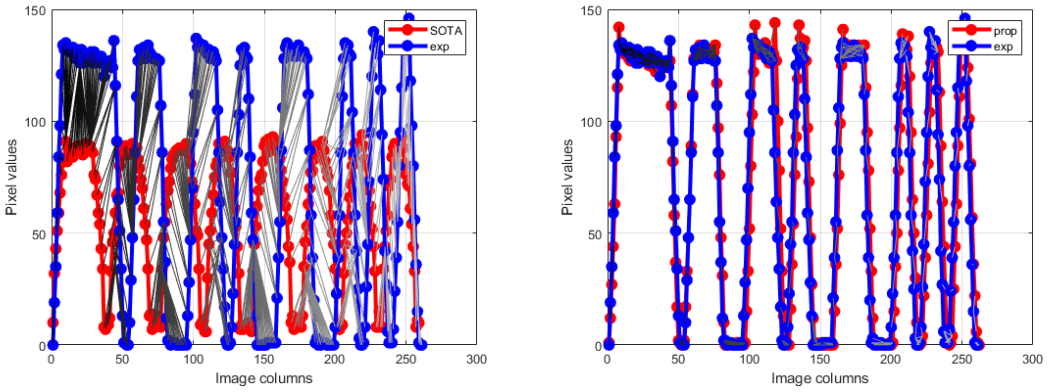
The pixel values when the exposure time is less than the switching period is shown in Fig. 5.3(a), where the three plots refer to the SOTA, the proposed technique and the experimental value for 10 kHz signal when the image was taken with an exposure time of 68μs 120 cm from the transmitter. The SOTA pixel values are lower than the other two since contrast enhancement was not performed. The logic used by SOTA to determine

complete and transition band lengths is shown to provide different pattern compared to the experimental value even when the exposure time is less than the switching period.

Since SOTA is not defined when exposure time is greater than the switching period, the pixel values for the proposed simulation technique and experimental values for are reported in Fig. 5.3(b). The pixel values of the proposed technique are marginally higher than the experimental values and the patterns are slightly different from each other. This can be due to the different image processing techniques used, but the overall difference between the two similar consecutive bits and one bit is apparent from both. Here the image was captured at the same settings with an exposure time of 136μs. As the exposure time increases the contrast between single ones and zeros reduces eventually changing the sequence of bits rendering detection impossible. The SOTA simulated images from the banding caused by the rolling shutter effect as a signal. This signal is calculated using the maximum and minimum pixel values along with the complete and transition band widths. The proposed technique uses the maximum and minimum pixel value calculations and calculates a weighted average based on the relation ship between the exposure time and switching period. This adds no further complexity to the computational algorithm of the SOTA technique. The difference in accuracy between the SOTA and proposed techniques is discussed in detail in the following section.

5.3.2 Discrete Fréchet Distance

While the similarity between the signal simulated through the proposed technique and the actual photo is apparent, we use discrete Fréchet distance[114] to quantify this similarity. This metric is the sum of the distance between corresponding points on two curves, which will be zero for the same curve and keep increasing as the curves become increasingly dissimilar. The three curves where the exposure time is less than the switching period in Fig. 5.3(a) are compared in pairs, the state of the art simulation technique and the proposed technique curves with the experimental curve. The curves generated by 10 bit OOK code at 10kHz switching speed with differential manchester encoding for 68μs exposure time are shown in Fig. 5.4. The figures show a clear picture of how the discrete Fréchet distance (DFD) is calculated. The lines which represent the Euclidean distance between the corresponding points in the curves become increasingly lighter from left to right. The SOTA simulated curve and the eperiental curve are compared in Fig. 5.4(a), where the difference in amplitude of pixel values contributes to the major difference in



(a) sota (b) proposed

FIGURE 5.4: Discrete Fréchet distance for points on the curve

the metric while the temporal difference is also visible though not as apparent. The curve simulated through the proposed technique and the experimental curve are compared in Fig. 5.4(b), where the lines between the curves are barely visible. There is amplitude and temporal differences are non existent compared to the state of the art technique. This difference is shown in the DFD of the two simulated curves being 13.43 for the proposed curve and 99.32 for the SOTA curve.

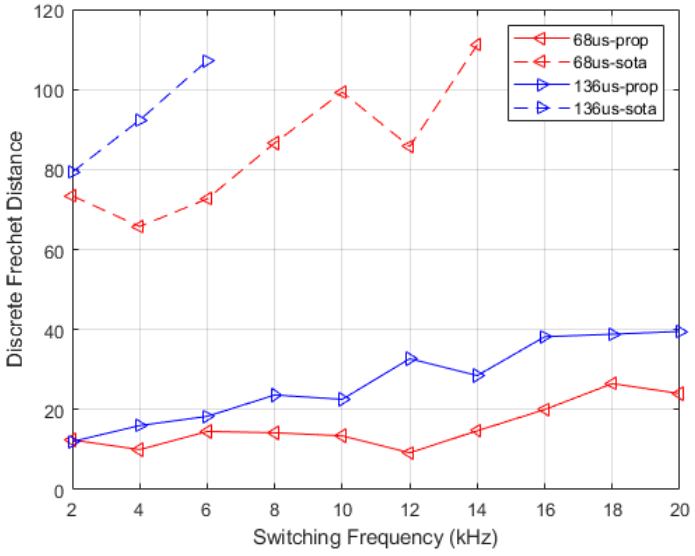


FIGURE 5.5: Discrete Fréchet distance as a function of switching frequency.

Though the difference in simulated curves is quantified for this specific switching period and exposure time combination, all other combinations need to be tested to draw wider conclusions on the simulation techniques. The DFD of the curves from the two

simulation techniques are plotted as a function of the switching frequencies in Fig. 5.5, where the exposure times are either 68μs or 136μs and the prop in legend refers to the proposed simulation technique while the sota refers to the state of the art simulation technique. At 68μs exposure time, the SOTA technique only has values until 14kHz since the switching period becomes 62.5μs which is less than the exposure time. The DFD of the SOTA curve and the experimental curve increases as the switching frequency increases with a couple of points decreasing slightly. This shows the equation 5.5 for complete and transition bands are not as accurate the proposed technique since the temporal mapping changes as the relationship between the switching period and exposure time changes. In case of the DFD between the proposed curve and the experimental curve for 68μs exposure time, the values are lower than 20 until 14 kHz switching frequency and increasing marginally beyond 20 when the switching period is less than the exposure time. For 68μs exposure time, the proposed technique has the highest DFD of 26.48, while the SOTA technique DFD was lowest at 65.74. Even when comparing the techniques within the frequency limitation of the SOTA technique, the DFD of SOTA is nearly three times the DFD of the proposed technique.

We further explored the difference in the simulation techniques when the exposure time was greater than the switching period as shown in Fig. 5.3(b) since the curve from the proposed simulation technique was shown to be increasingly dissimilar to the experimental curve. The DFD values at 136μs exposure time for proposed and SOTA techniques is shown in Fig. 5.5, where the SOTA curve only has 3 points till 6kHz since the switching period becomes 125μs which is lower than the exposure time. Even among these three points, the SOTA curve is shown to be dissimilar to the experimental curve with DFD increasing from 79.33 to 107.17. The DFD values for the proposed technique are calculated for all the frequencies ranging from 2 to 20kHz to show the influence of the difference between the switching period and exposure time on the proposed simulation technique. The DFD values till 6kHz are similar to those for 68μs exposure time both being less than 20, but the values increase as the difference between the switching period and exposure time increases with 39.52 DFD observed at 20kHz switching frequency when the switching period is 50μs and exposure time is 136μs. The worst DFD for the proposed technique is approximately half of the lowest SOTA DFD for the same exposure time. The curves shown in Fig. 5.3(b), at 10kHz switching frequency are noticeably dissimilar than those in Fig. 5.3(a) but the DFD is 22.55 showing the overall shape and amplitude remain consistent. Thus the curves from the proposed simulation technique was shown to be much more similar to the experimental curves than the curves from the

SOTA simulation technique. Though the similarity of curves from proposed simulation technique and the experimental curves decreases as the switching period decreases lower than the exposure time, the DFD was shown to be much lower than SOTA levels even when the switching period was less than half the exposure time. Thus, the performance of the proposed technique is shown to worsen as the exposure time increases with respect to the switching period. While this may be sufficient for the OOK modulation tested here owing to the existence of two extreme states, as the complexity of the modulation technique increases to incorporate dimming, these problems could be amplified at such large differences between the switching periods and exposure times.

5.4 Experimental Validation

5.4.1 Effect of transmitter parameters

Though we have shown the curves produced by the proposed simulation technique are more similar to the experimental curves than the SOTA simulation technique, we still do not know if these curves are similar enough to replace the experimental curves. Since we want to replace the experimental curves to help develop and test demodulation techniques we tested the simulated images on the conventional demodulation technique[110] and compared the resultant detection success rates with that from the experimental images. To account for the receiver parameter variation we have varied the switching frequency and exposure times but the transmitter variation has yet to be tested. We have incorporated a difference in the shape, size and luminous intensity of the transmitter panel in the image simulation process the specifications of which are outlined in table 5.1. A square panel light and circular panel light were tested and the results are reported.

5.4.1.1 Square Panel Light

The detection success rate is used to determine the accuracy of simulation and to ascertain the frequency at which detection stops. We have used two exposure times for both the devices tested, with one being the lowest possible exposure time for that camera and twice that value. The detection results for the phone are shown in Fig. 5.6, where the experimental success rate for both the exposure times is similar to the simulated values. The experimental success rate for both exposure times at 8kHz beyond 180cm is lower

than the simulated value but this is just a difference of two data points which could be chalked up to two headers not being observed for the experimental images. For the higher exposure time though the exact success rate is not observed, the technique provides a good indication of when detection ceases. This happens when the switching period is nearly half the exposure time at 14 kHz.

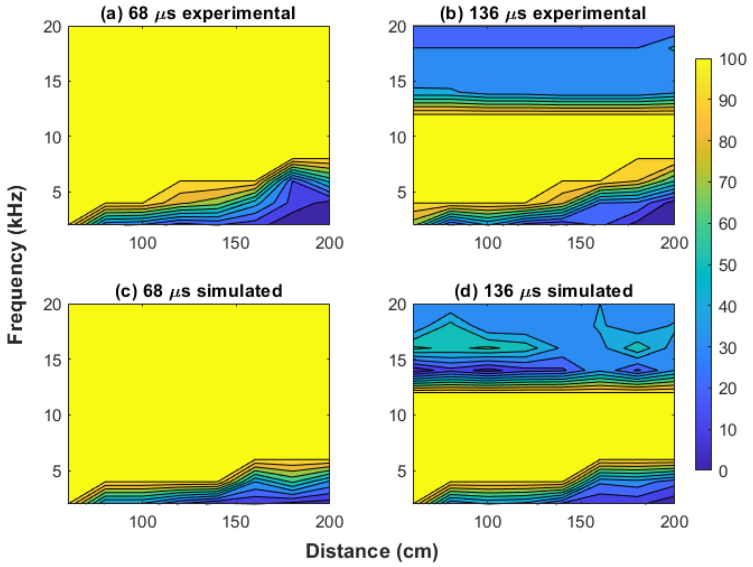


FIGURE 5.6: Detection success rate comparison for phone camera.

The success rates for the tablet camera are outlined in Fig. 5.7, where once again the experimental success rates are similar to the simulated values for both exposure times. The experimental success rate is lower than the simulated value at 2 kHz for both exposure times. We can determine that this frequency is low to accommodate an entire 10 bit sequence in the image. The detection ceases when the switching period is slightly greater than half the exposure time at 16 kHz. Thus, the proposed technique provides the ability to test detection techniques and to determine the ideal number of bits and switching frequency for a given exposure time.

5.4.1.2 Circular Panel Light

The detection success rate for circular panel light using the phone camera is shown in Fig. 5.8, where (a) and (b) are from experimental images while (d) and (c) are from simulated images. The simulated and experimental results at 68μs are similar. The switching frequency limit of 2 kHz is consistent across both the square and circular panel lights. The simulated and experimental results at 136μs are similar upto the 12 kHz beyond

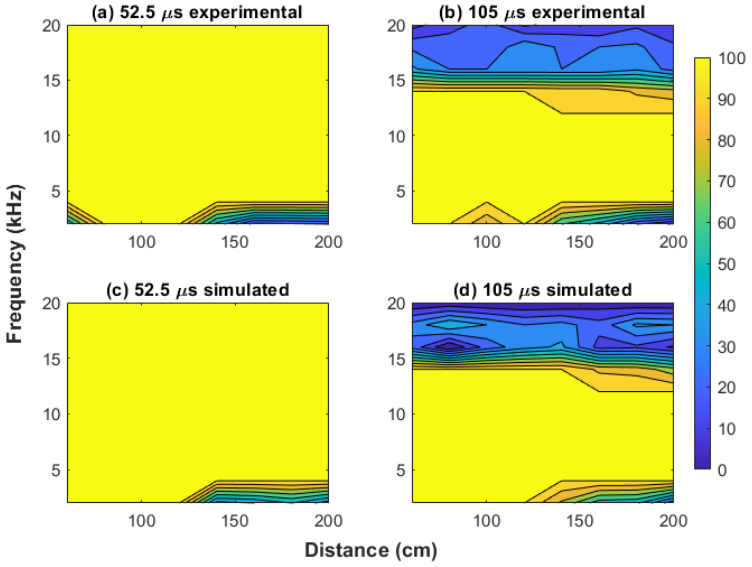


FIGURE 5.7: Detection success rate comparison for tablet camera.

which minor differences are observed but both remain undetectable with success rates predominantly below fifty percent.

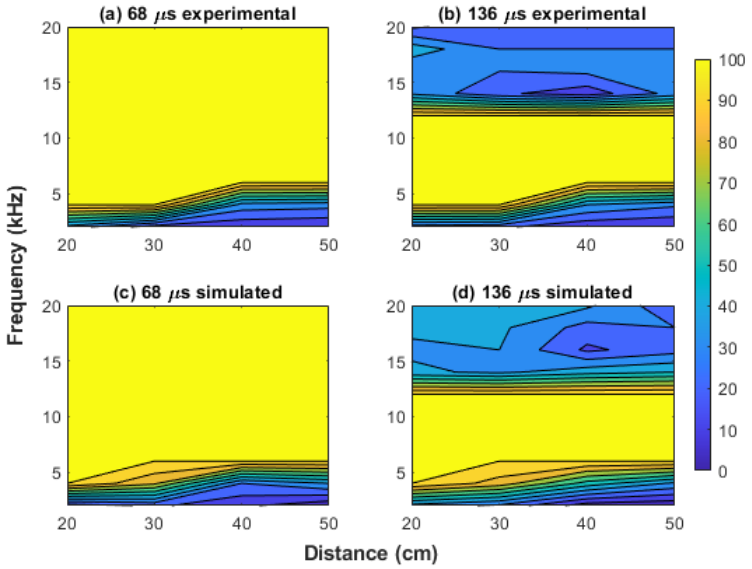


FIGURE 5.8: Detection success rate comparison for phone camera.

The detection success rate for a circular panel light using the tablet front camera is reported in Fig. 5.9, where even the minor differences observed for the phone camera was absent. The detection success rate at 52.5μs was the exact same value across all data points tested, with minor differences observed beyond the detection limit of 14 kHz in this case. The same detection limit was observed for both the devices establishing the

validity of the proposed simulation technique across different transmitter and receiver specifications.

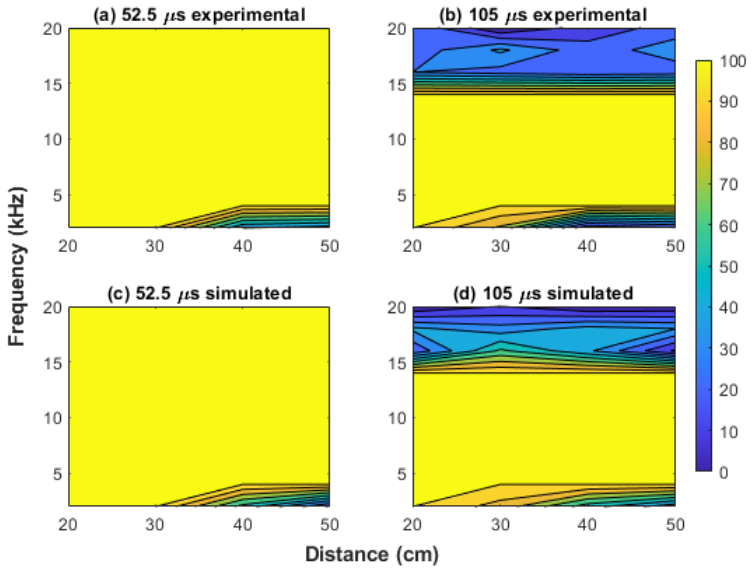


FIGURE 5.9: Detection success rate comparison for tablet camera.

5.4.2 Effect of Noise

The images from the proposed simulation technique was shown to be similar enough to replace experimental images for demodulation algorithm development and testing. The simulation techniques also provide a way for us to test when an algorithm will breakdown as the noise in the image increases. To test this on the conventional technique, we added Gaussian noise to the simulated pixel value at a specified distance from the transmitter to observe its effects. We tested the effect of noise on both the phone and tablet front cameras at 136μs and 105μs exposure times respectively since these were shown to have a switching frequency limit beyond which the conventional technique fails. Apart from the receiver variation, the transmitter variation has also been incorporated into the testing with both the circular and square panel lights. The success rate is plotted as a function of switching frequency and noise level in Fig. 5.10, where each noise level indicates an increase of 10 standard deviation to the Gaussian noise.

For the square panel light, the phone and tablet camera images are Fig. 5.10(a) and Fig. 5.10(b) respectively. The results are aggregated at 1m from the transmitter for the square panel light. The phone camera at 136μs showed that beyond the 12kHz switching frequency the conventional technique fails as shown in Fig. 5.6. This is observed in

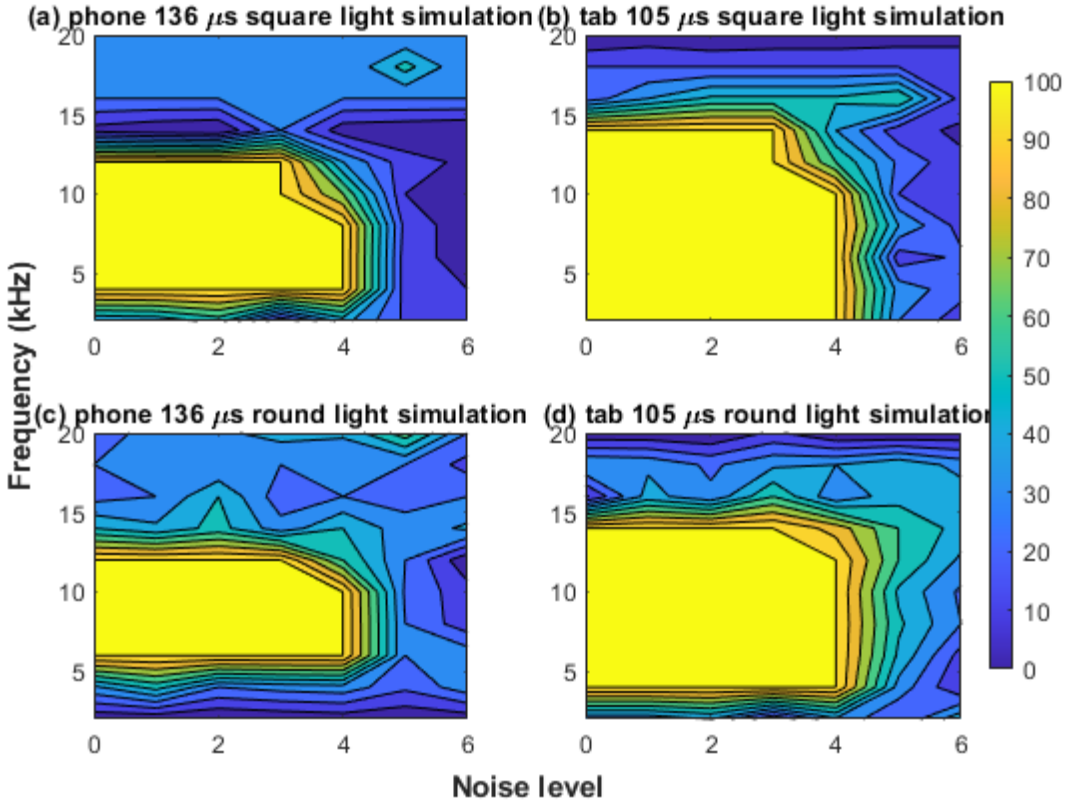
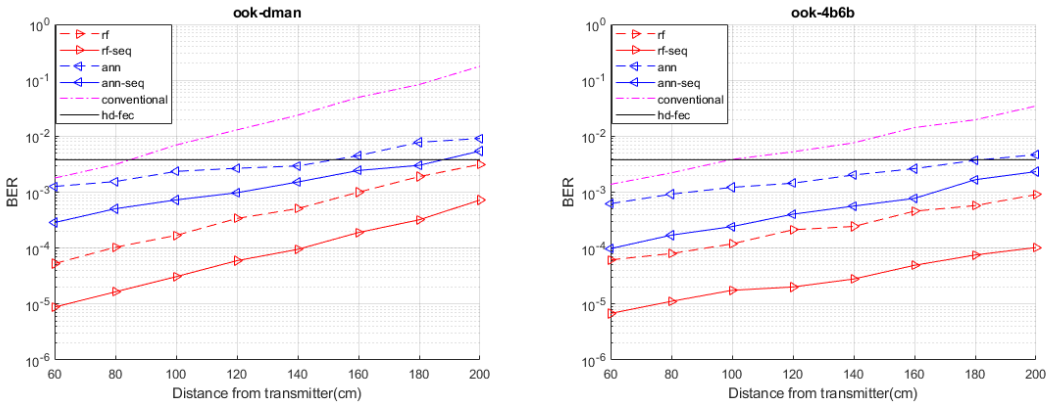


FIGURE 5.10: Influence of noise on the detection success rate.

Fig. 5.10(a) when the noise level is zero. As the noise level increases we see no major differences until the noise level three and four, where the detection fails at 10kHz for three and 8kHz for four which shows the increased sensitivity of the decoding technique to noise. A similar transition is observed at noise level three and four in Fig. 5.10(b), where the switching frequency limit is 14kHz for the tablet camera and a reduction 12kHz and 10kHz was observed. The detection success rates were aggregated at 0.5m for the circular panel light owing to its smaller size compared to the square panel light.

The detection limit changes to 10kHz from 12kHz for the phone camera in Fig. 5.10(c) and from 14kHz to 12kHz for the tablet camera in Fig. 5.10(d) when the noise level is four. There is a two step reduction for the square panel light consistent across both receivers while only a single step reduction is observed in the circular panel light. Beyond noise level four no transmitter and receiver combination produces detectable signals. The difference in resolution of the two cameras tested was also apparent from the success rates reported. The phone camera has a lower resolution which makes detection impossible until 4kHz for the square panel light and 6kHz for the circular light. The tablet camera has a higher resolution which makes more states of the light visible making detection



(a) manchester encoding (b) 4b6b encoding

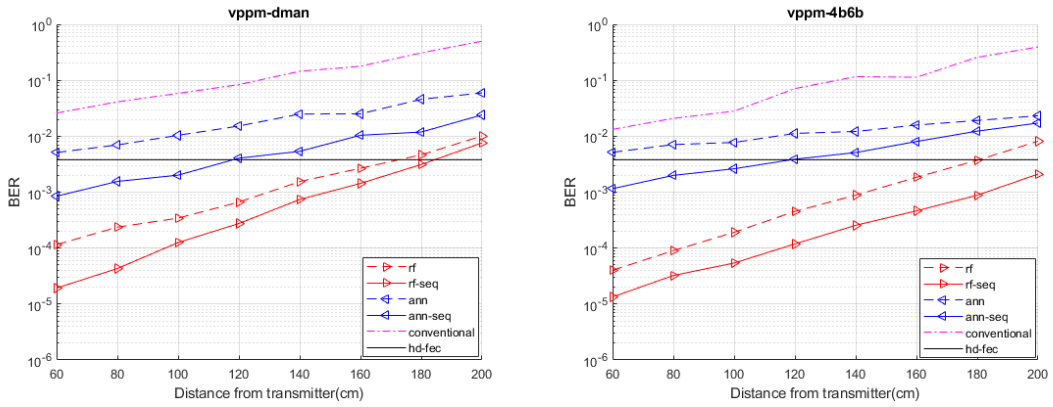
FIGURE 5.11: Bit Error Rate at different distances from the transmitter for OOK

possible at lower frequencies of 2kHz for the square light and 4kHz for the circular light. Thus the proposed simulation technique was shown to be capable of testing the effect of noise on detection techniques.

5.4.3 Effect of modulation and encoding

While the proposed simulation technique was experimentally validated using the detection success rate, to test the effect of modulation and encoding on the simulation technique we reported the bit error rate (BER). Since a larger range of frequencies, transmitter and receiver parameter were tested to validate the simulation technique detection success rate which was reported using few images was employed. Since BER requires a large number of frames to draw effective conclusions, only two modulation and two encoding techniques were tested using the phone camera and square panel light. We captured a one minute video at 60 frames per second (FPS) in 1920X1080 resolution. This video was then broken down into 3600 frames from which the BER results are reported. Apart from the OOK modulation and manchester encoding techniques, we tested commonly used variable pulse position modulation (VPPM) and 4b6b encoding techniques.

The BER results for OOK modulation are outlined in Fig. 5.11, where BER is plotted as a function of distance from the transmitter. The results for manchester encoding are shown in Fig. 5.11(a), where ann denotes an artificial neural network and rf denotes random forest the suffix -seq denotes the models trained with the output of the previous bit as an additional feature. The results from conventional technique are also compared



(a) manchester encoding (b) 4b6b encoding

FIGURE 5.12: Bit Error Rate at different distances from the transmitter for VPPM

with the same. The hd-fec refers to the hard decode BER limit before forward error correction which is 3.8×10^{-3}. The BER is reported from 60cm to 200cm in 20cm increments where the BER was observed to increase linearly with increasing distance from the transmitter. A significant improvement in BER was observed between the ann and rf models, with slight improvements introduced by the previous output bit feature. The worst performing among these models was still observed to produce better BER at all distances. The results for 4b6b encoding were shown in Fig. 5.11(b), where all the models performed marginally better than the manchester encoded data owing to the fewer number of bits required to represent the same data. Similar patterns were observed with the proposed improvement to the demodulation technique producing better BER compared to both the conventional technique and the basic models.

The BER results for VPPM modulation are delineated in Fig. 5.12, where the results are worse than those from Fig. 5.11 due to the higher number of bits required to represent the same data. Among the two proposed models ann performs worse than the tree-based rf model. The ann model performed worse than the hd-fec level at 2m from the transmitter for manchester encoding and never for the 4b6b coding in the case of OOK modulation. For VPPM with manchester encoding the ann model performs worse than the hd-fec level at 1.2m from the transmitter and the rf model breaches this level at 2m from the transmitter as shown in Fig. 5.12(a). The combination of manchester encoding with VPPM produces the worst results overall owing to requiring the most number of bits to represent the same eight bit data. The VPPM with 4b6b condition is reported in Fig. 5.12(b), where the performance is better than Fig. 5.12(a) but worse than Fig. 5.11. The rf model here manages to stay below the hd-fec level at all distances but the ann

model starts performing worse than the level at 1.4m from the transmitter. Thus we can conclude OOK modulation performs better than VPPM and 4b6b encoding is better than manchester encoding. The proposed demodulation technique was shown to perform better than the conventional technique and models without the additional feature. This technique does away with the influence of the background features on the performance by operating at the lowest possible exposure time. However, if the transmitter is placed too close to a more powerful light source or if the light is placed in the corridor or balcony where external light can interfere with the signal the proposed technique and trained model will no longer perform as reported but this would also influence the conventional technique used in this section.

5.5 Conclusion

A simple simulation technique for camera based VLC at any signal frequency irrespective of exposure time was proposed. The simulated images were compared with experimental images. The accuracy of the simulation technique was tested using a simple detection technique to compare detection success rates of simulated with actual images. Thorough experimental validation was conducted using two different devices at two exposure times for a range of switching frequencies and distances. The simulated and experimental success rates where shown to be similar up to the switching frequency where detection was no longer possible for both the devices. Different modulation schemes and coding techniques were further explored by analysing BER of a proposed demodulation technique with the conventional demodulation technique.

While there are more advanced modulation techniques capable of producing higher data rates, the simpler techniques were used since the aim is to facilitate coarse indoor positioning using OCC. The more complicated techniques will need dedicated hardware such as specialised transmitters and better cameras or photodetectors on the receiver for demodulation. We have explored these simple techniques with COTS transmitters and cheap Android phones to ensure an indoor positioning system can be implemented with what is available now. The proposed simulation technique ensures data for developing demodulation techniques and choosing transmitters and optimising their placement can be obtained easily without having to procure and install lights. The proposed simulation and demodulation technique aid in the coarse indoor localisation as part of the overall indoor positioning problem.

The proposed technique is important because it shows the limit of implementation using commercial off the shelf lights and smartphone cameras. This technique can be used for indoor positioning as a part of many location based services (LBS). In museums and exhibitions providing more information about a specific exhibit can be expensive requiring individual display screens for each exhibit. This can be solved by providing this information through a video or image format as an LBS. A simple content delivery network on the back-end of a mobile application can allow for videos about each exhibit to be viewed by the user on their smartphones. In hospitals, where information about the patients is written on notepads at the foot of their beds to apprise doctors of their status this technique can hide the data on a server away from the prying eyes of passers-by. The improved positioning accuracy and improved navigation of visitors in restricted areas such as offices offers two fold benefits. The visitor can be given granular directions to the desk of the person they are visiting and the accurate location of the visitor can be tracked to ensure they do not visit any restricted section of the office space.

Chapter 6

WiFi based coarse indoor positioning

6.1 Introduction

Extant indoor positioning techniques based on received signal strength (RSS) fingerprinting have achieved accurate positioning results. They leverage existing open-source datasets for training and comparing their model performance. However, the models trained on one building cannot be used for another since these models learn the relationship of a specific set of RSSs to the building and floor locations necessitating the expensive, time-consuming process of fingerprinting. Even when we consider the individual datasets producing these excellent results, a lot of painstaking optimization is required which precludes a lot of people trying to implement indoor positioning quickly. Most of the input RSS vector is empty with redundant information and the static class labels used for buildings and floors make the models unusable on other buildings. This chapter proposes a machine learning-based framework that uses RSS values from the strongest access point (AP) signals and normalized output labels to combat this issue. The framework was used on the open-source UJI dataset using less than 5% of the 520 APs to achieve 94.15% and 8.45 m floor prediction accuracy and mean positioning error respectively without any optimization. This technique was reused on 10 other public datasets and achieved an average floor estimation accuracy of 91.93% when trained with new data and 88.68% without any new data compared to 87.1% of the closest competitor.

All previous chapters discuss the visible light positioning problem which works when the user is present under a light such that it is captured in the field of view of the camera. Though lights are present throughout the building there will be several points where

the lights are not visible or partially visible in the field of view of the receiver. This is due to the lights currently being installed sparsely for illumination alone. When we are between lights, we propose using Wi-Fi fingerprinting to limit outage. Owing to increased interest in location based services such as indoor navigation[74] and targeted advertising[115], we seek to produce a transferable indoor positioning model for WiFi based fingerprinting. While GPS fails for indoor positioning[1], Bluetooth[8], radio frequency identification (RFID)[7], ultra-wideband (UWB)[17] and visible light based positioning[85] require expensive infrastructure deployment to achieve high accuracy. Received signal strength (RSS) and fingerprinting are popular since it allows for decimetre level positioning accuracy using smartphones and existing WiFi infrastructure[29]. While CapsLoc[116] achieved high accuracy using capsule networks in the fixed input output structure currently used for WiFi fingerprinting, in order to have the model learn a more general representation of the problem and do away with the large input vectors that require autoencoders or other computationally intensive methods for representation learning[68, 117, 118], we changed the input and output structure while using the same public datasets to ensure the results produced here are comparable to similar techniques. To avoid confusion arising from APs not being assigned to the same input node the location information of APs with respect to the strongest AP was also included as the input. We solved an important part of the problem since feature extraction before classification or regression produces better results [29]. This reduction in input features led to a reduction in training time, size and complexity of the model, which is important[119] since most WiFi fingerprinting models are deployed on limited hardware such as smart phones and in conjunction with other techniques such as pedestrian dead reckoning[120] or visible light positioning[80].

The concept of a transferable model has been tried several times, with a common deep learning core and an autoencoder to account for different input lengths [117] or normalizing the number of floors in a building to get closer floor estimates [121] but none managed to achieve the goal as well owing to their failure to tackle the input and output relationship problem. There have been other models where normalization techniques were employed to improve generalizability. The RSS values were normalized based on AP distances to obtain an image which was then used to train a convolutional neural network(CNN) classifier for floor and building estimation [122] and produced high classification accuracy but cannot be transferred to other datasets since the resolution of the image restricts the dataset size to be above a certain number. While MetaLoc [123] has proven capable of achieving better accuracy it still requires fingerprinting data and retraining to achieve

those results. The problem with these complicated deep learning models is that the reported results are achieved after painstaking optimization. Doing away with the need for optimization makes it easier to achieve accurate results quickly, which is essential in real-life deployment of indoor positioning systems.

The main contributions of this work are listed as follows

- A simple series of steps anyone without domain expertise,in data analysis, can follow to achieve high accuracy out of the box was proposed.
- A new dimensionality reduction technique was proposed to retain the important information from the sparse RSS vector and was compared to current standard techniques.
- New input and output normalization schemes were presented to facilitate transferability of the proposed technique.
- The performance of the proposed technique with estimated AP location was compared with the accuracy achieved using the actual AP location.
- The generalizability of the proposed technique was established by rigorous testing on 11 datasets.

6.2 Methodology

6.2.1 Proposed structure

The first step in the proposed technique is identifying the AP locations using an AP localization strategy as shown in Fig. 6.1. Feature extraction here refers to the sorting of RSS based on signal strength and finding the difference in AP locations compared to the strongest AP. This input dataset is then split into the training, validation and test datasets and used to train the machine learning model after which we obtain the trained model and its corresponding accuracy metrics from the test dataset.This trained model is then implemented in a new building with RSS data from a previously unseen mobile phone (receiver) and a different number of APs organized in a different structure. Since we have used other public datasets apart from the one used for training, we had to estimate

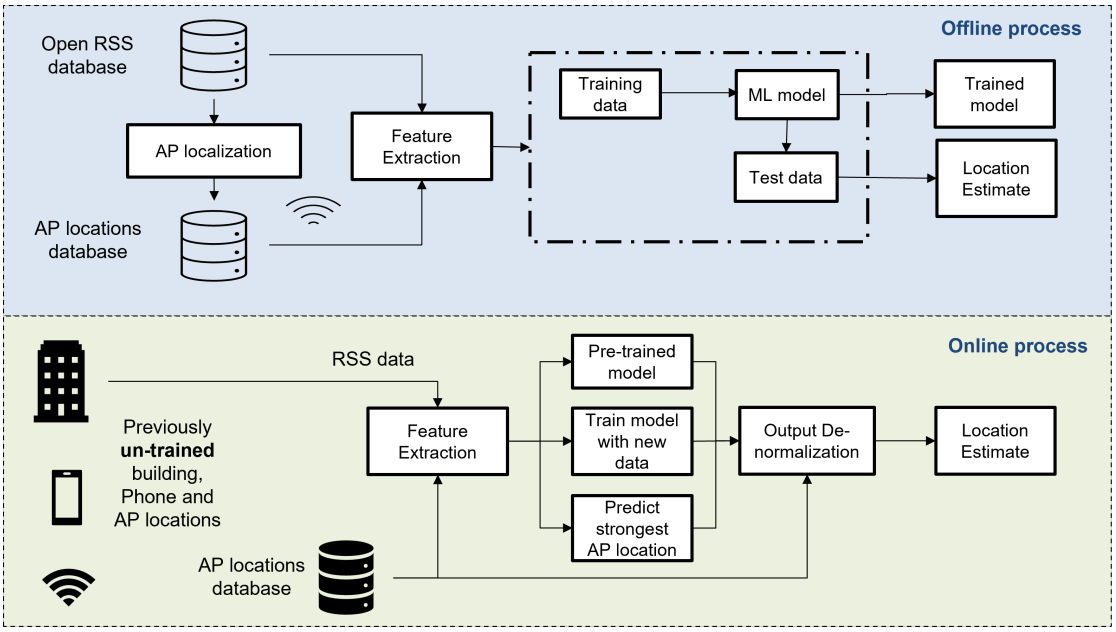


FIGURE 6.1: Proposed Structure.

AP location but, in most cases, indoor navigation or location-based services in commercial spaces such as office buildings, malls or university campuses are implemented by building owners who have all the AP location information.

The three stages after feature extraction are used to test the impact of the different steps in the proposed pipeline. To test the performance of the dimensionality reduction technique implemented in the feature extraction block, the strongest AP location is predicted to be the location of the RP in question. To test the effectiveness of the pre-trained model, the model trained on the UJI dataset[74] was used to predict the location estimate for 10 other publicly available datasets. To test the overall framework being proposed here, the available datasets were trained using the same technique used in the offline process. This gives us a normalized estimate of the location which is then denormalized using the AP location information to produce actual location estimates.

6.2.2 Transmitter localization

The public datasets generally do not contain AP location, which is why we need to perform the localization. However, to decide which technique performs best we need ground truth since there are several techniques to choose from. The LIB1 dataset [124] has the AP locations for 16 of its 174 APs and these are located on two different floors. The 16

APs are recorded from 4 devices as shown in Fig. 6.2. The RPs, where the data was collected from, are evenly distributed over the area and the APs lie on the extremes flanked by data points which makes it ideal for the test case. Since this dataset was collected over fifteen months it has a hundred thousand data points for each AP. The following subsections will cover the different techniques tested, the results and the final choice for subsequent testing.

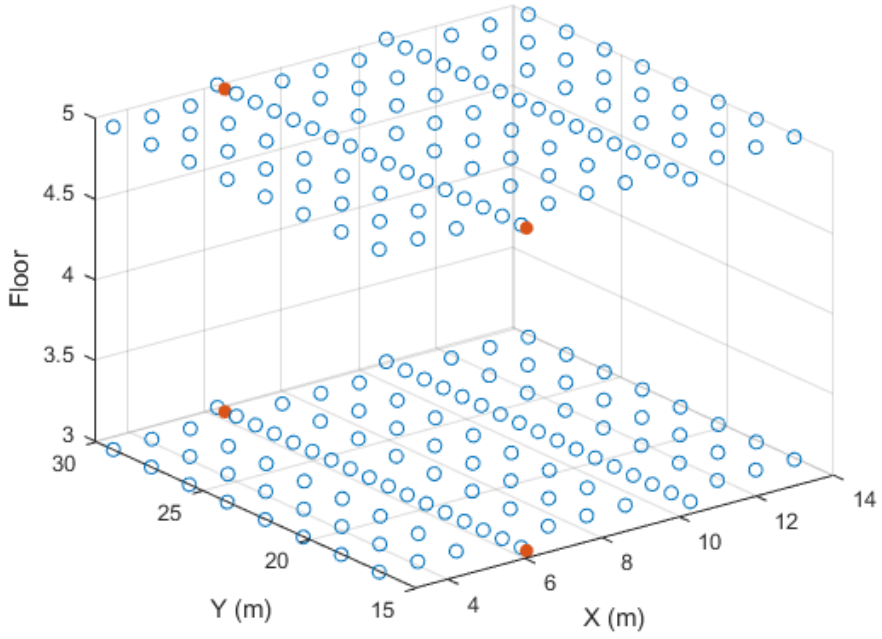


FIGURE 6.2: LIB1 dataset, blue dots are RPs and red dots are APs.

6.2.2.1 Techniques

Six different techniques were tested for AP localization. The simplest strategy employed was to assign the RP location with the highest signal strength as the transmitter location. In case there are multiple points with the same maximum value the first RP instance of the value was chosen to be the AP location. The second technique was another relatively simple technique called normal centroid[125] shown below

$$\hat{\mathbf{m}}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} \mathbf{p}_i^j \tag{6.1}$$

where, $\hat{\mathbf{m}}_j$ is the estimated location of the jth transmitter, N_j is the number of RPs the jth AP was heard in and $\mathbf{p}_i^j$ is the RP location the jth AP was heard in. This however does not take the signal strength into consideration, so two variations of the weighted centroid method[126, 127], which was originally proposed for wireless sensor nodes, were tested, the general equation of which is given below

$$\hat{\mathbf{m}}_j = \frac{1}{\sum_{i=1}^{N_j} w\left(r_{ij}\right)} \sum_{i=1}^{N_j} w\left(r_{ij}\right) \cdot \mathbf{p}_i^j \tag{6.2}$$

where, r_{ij} is the RSS of the jth transmitter recorded at the ith RP and $w\left(r_{ij}\right)$ is the weight assigned to r_{ij}. Instead of taking the average of all RP locations a particular AP was heard in, this introduces a weight for each location based on the signal strength. There are two different ways in which this was implemented. The first being distance based weighted centroid[128], the weight for which is calculated as follows

$$w\left(r_{ij}\right) = 10^{\lambda r_{ij}} \tag{6.3}$$

The second is the RSS based weighted centroid[127] which is calculated as follows

$$w\left(r_{ij}\right) = \left(r_{ij} - r_{\min}\right)^{\lambda} \tag{6.4}$$

where, $r_{\min}$ is chosen to be a value much lower than the minimum detected RSS value for the dataset. λ was chosen to be 0.07 for distance-based and 5 for RSS-based techniques based on genetic algorithm based optimisation performed by Nurminen et.al.[129]. Till now simple, computationally light techniques were tested. Now more complicated and accurate techniques are to be tested. The log-distance path loss (LDPL) model[130], given below, is a widely used approximation for any wireless communication system.

$$r(d) = r(d_{(0)}) - 10n \log_{10}\left(\frac{d}{d_{(0)}}\right) + x_\sigma \tag{6.5}$$

where, $r(d)$ is the rss at a distance d from the AP, d_0 is a reference distance from the AP for which the RSS should be known, we have chosen it to be one meter, n is the path loss exponent which shows the rate at which the RSS deteriorates with increase in distance from the AP and x_σ represents a normal distribution with standard deviation σ. The remaining two techniques focus on solving a non linear least square problem which

is detailed below

$$\left(\widehat{\mathbf{m}}_j, \widehat{\theta}\right) = \arg\min_{(\mathbf{m},\theta)} \sum_{i=1}^{N} \sum_{i} \left(h_i(\mathbf{m}, \theta)\right)^2 \tag{6.6}$$

where $\widehat{\theta}$ is the set of all parameters apart from AP location the optimization scheme will solve for and $h_i(\mathbf{m}, \theta)$ is the cost function used for minimization which is shown below

$$h_i(\mathbf{m}, \theta) = r_{ij} - r_{(0)}(\theta) + 10n(\theta) \log_{10} \left\|\mathbf{p}_i - \mathbf{m}\right\| \tag{6.7}$$

The two techniques used for solving this optimization problem are Levenberg-Marquardt(LM) which is a type of Gauss Newton optimization[131] and non-dominated sorting genetic algorithm, NSGA-II[132]. The LM technique was employed for transmitter localization in [133, 134]. These techniques will give us the results required for estimating x, y and z coordinates. However, if there are multiple buildings as is the case in UJI as shown in

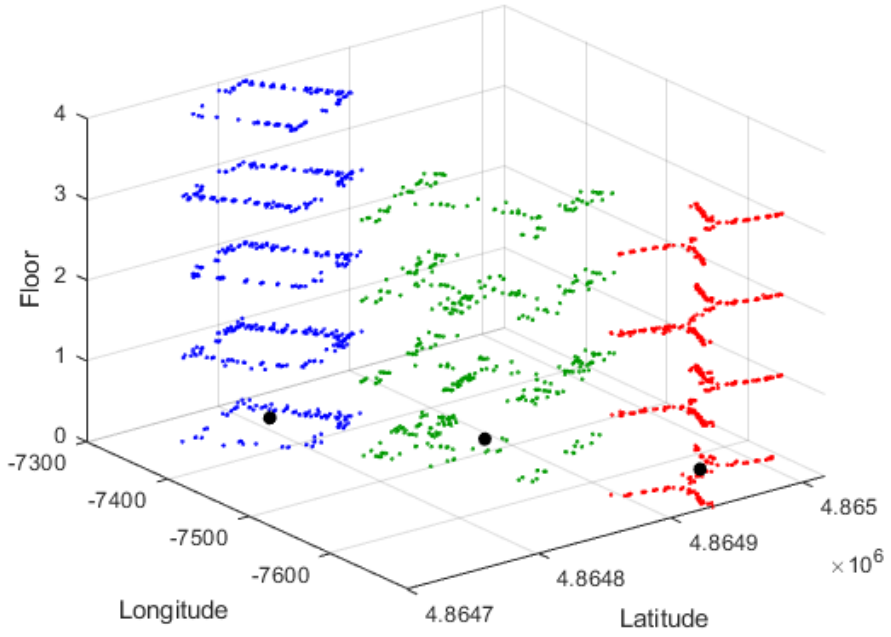


FIGURE 6.3: UJI dataset, black dots are the centres of buildings.

Fig. 6.3, where the different buildings are denoted in different colours, then we use the center of each building marked with black dots in Fig. 6.3. The distance of an estimated AP location from the center of the three buildings is calculated and the building with the lowest distance is chosen to be the building of the transmitter.

6.2.2.2 Results

The results observed from the aforementioned six techniques are shown in Fig. 6.4. The normal centroid based localization technique is labelled C in the figure, which performs the worst of the six techniques tested with a median mean square error(MSE) close to seven meters. The second worst is the RSS based weighted centroid method, marked WC_ RSS in the figure, which produces close to four meters MSE. This is followed by the distance based weighted centroid, marked WC_ DIST, which has a median MSE of nearly three meters. The simple strength based technique, labelled MS, outperforms the closed-form equations with an approximately two meter median MSE though with a couple of outliers beyond the five meter range. The more complicated optimization techniques expectedly outperform all other techniques with sub-meter median MSEs. NSGA-II produced slightly better accuracy compared to Levenberg-Marquardt, marked GN in Fig. 6.4, but has one outlier with greater than three meters MSE.

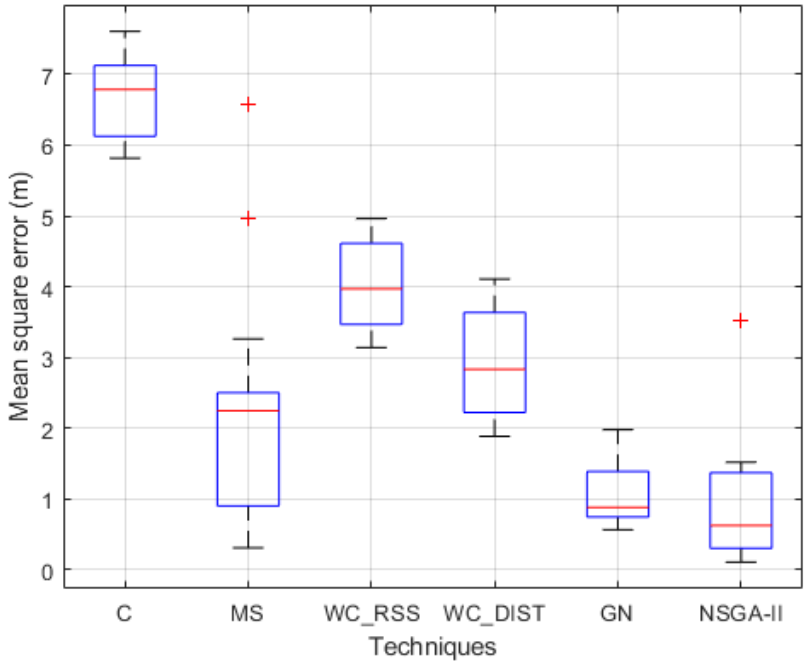


FIGURE 6.4: Mean square error of different AP localization techniques.

6.2.2.3 Effect of data

From the AP localization results obtained it might be tempting to choose the computationally heavy optimization techniques for localization. However, since these buildings

span hundreds of meters and multiple floors anything below five meters MSE would be a close enough estimate for the use case. Moreover, we have used a small subset of the LIB1 dataset which has the ground truth for AP location and this dataset has several thousand points. This is not the case for all other datasets since some are crowd-sourced and few are collected over such long periods of time with such granularity.

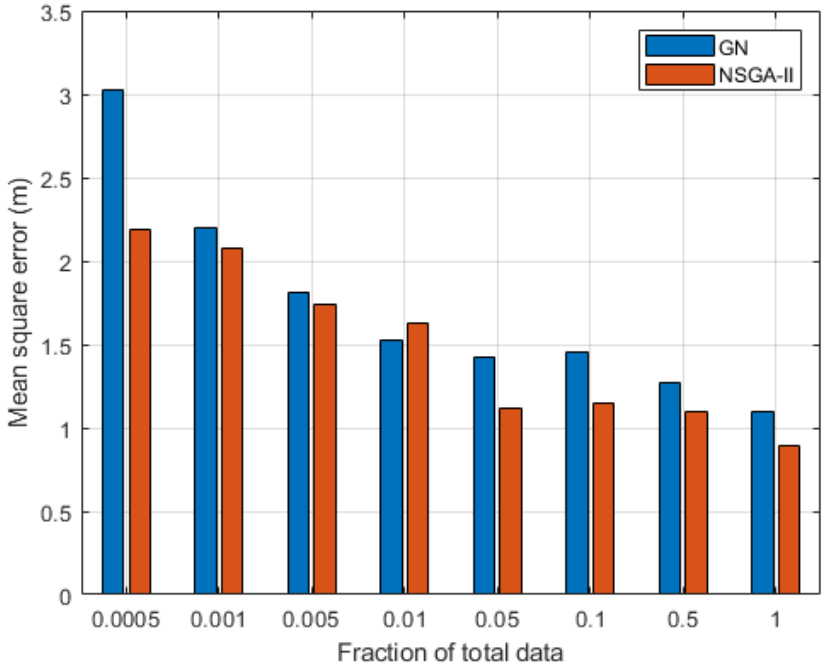


FIGURE 6.5: Effect of data on mean square error.

As the amount of data available for training reduces the performance of these optimization techniques also decreases commensurately as seen in Fig. 6.5. Although the MSE plateaus from 1% to 10% of the data for both Levenberg-Marquardt, labelled GN in figure, and NSGA-II, when the amount of data is reduced to around 50 points the result is in the vicinity of other techniques in Fig. 6.4. The distribution of the number of APs heard at less than 100 RPs in the UJI dataset is shown in Fig. 6.6. We can see that 107 of the 520 APs were heard at less than 10 RPs which makes them unusable. Hence we have chosen to proceed with the maximum RSS and both the weighted centroid techniques for further testing.

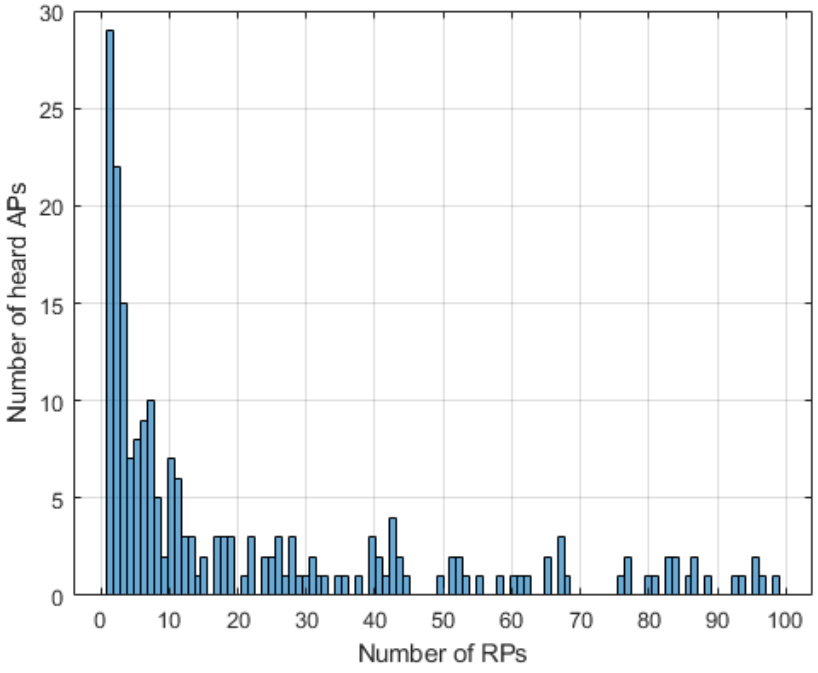


FIGURE 6.6: UJI data RP distribution.

6.2.3 Normalization

6.2.3.1 Input dimensionality reduction

Two major areas need to be addressed for normalization, which are the inputs and outputs. The structure of the normalized model is shown in Fig. 6.7, where each node in the input side represents w input features, where w is the number of APs. The number of APs determines the overall length of the input vector. The AP location was used to normalize the input features across multiple datasets. Even among the datasets being tested in this chapter the RP location, which was used for AP location estimation, is not provided in the same format. UJI has latitude and longitude coordinates in meters with UTM from WGS84 [74], while the datasets that have coordinates in meters also have the origin at different locations. To correct for these differences, the centroid of the region covered by the dataset was made the origin for all the datasets that were tested by subtracting the centroid from all the RP locations. This ensured the AP locations being calculated were comparable between different datasets. The RSS_w node represents the strongest w input RSS values in descending order followed by ΔX_w to represent the vector of x position difference between the strongest and other $w-1$ nodes.

$$\Delta x_j = x_j - x_1 \tag{6.8}$$

where, Δx_j is the individual x value difference in ΔX_w and j takes values from 2 to the number of APs w. x_1 is the x-coordinate of the strongest AP location since the values were sorted in descending order. The vector ΔX_w contains $(w-1)$ values. Similarly, the difference was calculated for the transmitter y coordinate, building, floor, and the distance between the strongest and other APs represented by ΔY_w, ΔB_w, ΔF_w and Δd_w respectively. Hence, the length of inputs is $w + (5(w-1))$ which makes the choice of w important to determine the complexity of the model required to learn the relationship. In the case of the 11 datasets tested, the average number of APs heard at each RP was 15, however, the floor hit rate and mean positioning error were found to be better than the state of the art at five and ten APs respectively. Hence showing that the average number of heard APs can be a good starting point considering the hundreds of APs in each dataset.

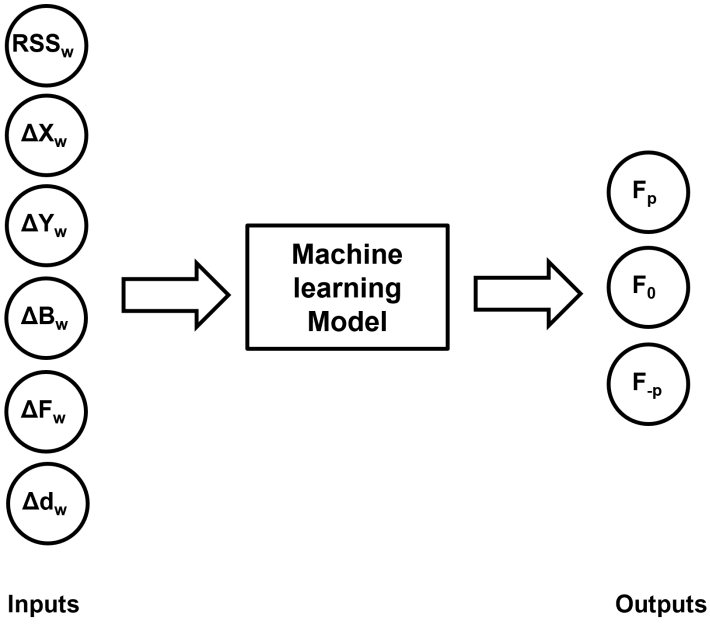


FIGURE 6.7: Normalized input output structure for floor classifier.

6.2.3.2 Output normalization

The second part of the normalization is in the output end. To estimate the floor a classifier will be trained since there are applications where just the floor level localization

is required.Depending on the number of floors or buildings being tested we can increase or decrease the number of nodes. However, having more than the number of classes required will neither hamper performance nor need retraining but having fewer classes than required will require the model to be retrained to incorporate data for the new classes.

As we can see the normalization is done with respect to the strongest AP location as in the case of the input nodes. We find the difference between the actual RP location, normalized as outlined in the previous section to have its origin at the centroid of the location covered by the dataset in question, and the strongest AP location at that RP. Similarly, the x and y coordinates are predicted as the x and y difference between the strongest AP location and the RP location. This gives us class labels and locations that allow the model to learn the relationship between RP and AP locations with respect to each other. This allows us to use the same model without retraining on a building from a different part of the world with no loss in accuracy. Once we get the normalized floor and location estimate, we can add the floor or x and y location of the strongest AP to get back the actual estimates, this is called output de-normalization in Fig. 6.1.

The output nodes in Fig. 6.7 are F_x where x represents the difference between the strongest AP location at an RP and the actual RP location. When x is negative, the actual location is below the strongest AP location while a positive value means its above the strongest AP location. The p in f_p represents the difference in floors we want our model to accommodate. While it is easy to choose twice the total number of floors, we have chosen 9 which is one more than a four floor difference on either side. This was chosen since we seldom encounter APs producing RSS more than four floors apart high enough to be classed among the strongest 5 or 10 APs. The same process is repeated for the x and y location but since the difference is a continuous value, instead of a classification problem it becomes a regression problem which does not restrict the outputs to a fixed number of labels. This type of output normalization was tried with a different input structure which produced much lower prediction accuracy[135].

6.3 Results and Discussion

6.3.1 Effect of transmitter localization technique

Three different transmitter localization techniques were tested to determine the effect of AP location estimation on the localization performance. The simplest strategy employed was to assign the RP location with the highest signal strength as the transmitter location. Weighted centroid techniques were also tested owing to their simple closed form nature. The distance based[128] and RSS based[126] weighted centroid were the other two techniques tested. The free parameter in both these techniques were chosen to be 0.07 for distance-based and 5 for RSS-based techniques based on genetic algorithm based optimisation performed by Nurminen et.al.[129]. A dataset which covered three consecutive floors with the ground truth location of APs was used. The Hybrid-fingerprint Data with Layout Change(HDLC) dataset [136] as the name suggests was collected to show the difference a layout change produced in the positioning results. This dataset also had BLE data apart from WiFi, but we tested the proposed technique on the three different layouts with the different transmitter localization techniques and compared the positioning results with those obtained when the actual AP location was used instead. The HDLC dataset contained three layouts which covered the same corridors in a three floor Faculty of Engineering building at the Multimedia University in Malaysia. The layouts were progressively obstructed by boards to simulate construction or furniture being moved around a building. Each layout was split into a train and test dataset by the provider, and the same split was used for training the proposed technique. The mean 3D positioning error, which was the average Euclidean distance between the predicted location and the actual location, is shown in Fig. 6.8, which contains three graphs. The training dataset used to train the random forest regression model is indicated in the title and results are averaged over the three test layout datasets. The default values from scikit-learn[99] for random forest and other models were used in all the cases.

There were three subplots, for the three layouts in Fig. 6.8 each of which in turn had three line plots, which were the mean positioning errors obtained when we used three different types of AP locations to calculate the input features. The first of these was labelled actual, where the actual AP position was used. The line labelled average used the average of the estimated AP locations from the three transmitter localization techniques used here and the line labelled min used the estimated AP location which produced the minimum positioning error value among the three transmitter localization techniques tested in this

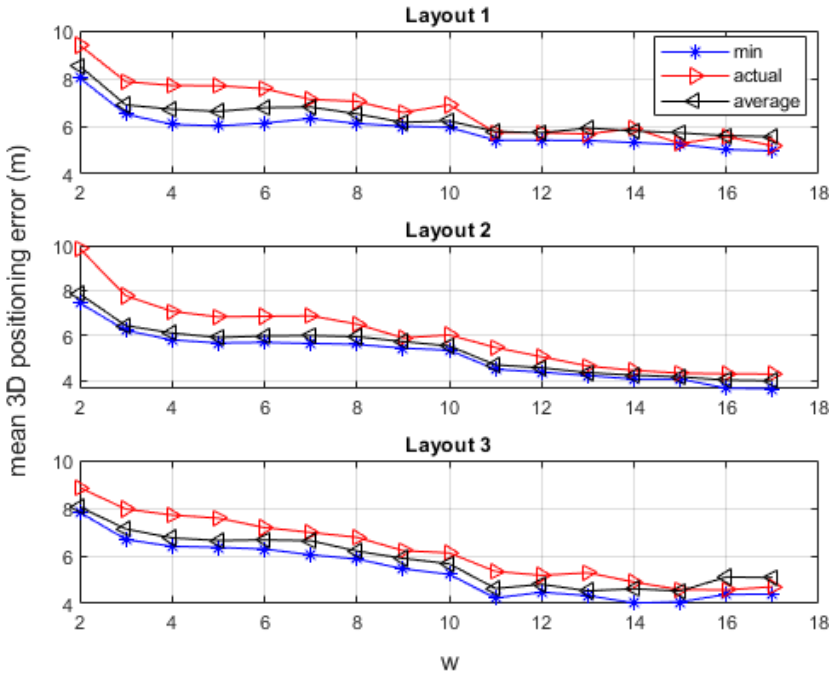


FIGURE 6.8: Mean 3D positioning error for the HDLC dataset.

chapter. Since the dataset covers 17 APs, the w values, the number of APs, in the x axis range from 2 to 17. We can see from Fig. 6.8 that between the average of the three transmitter localization techniques and the actual AP location for lower number of APs, the latter trails the former but when the number of APs was closer to the total number available, there was very little between the techniques. The best performing technique for each w consistently outperforms both the other techniques. In the case of layout 2, a transitory number of obstructions between layout 1 which had less and layout 3 which had more, the performance of the average was marginally better than the actual location but in both the other cases owing to layout 3 and 1 being very different from each other the actual position performed better than the transmitter localization technique average.

To study the performance of the proposed technique for applications that only need floor level localization, we trained a random forest classifier on the three different training datasets from the three layouts and the floor hit rate was graphed for the three techniques as shown in Fig. 6.9. The labels of the line plots refer to the same types of AP locations as used in Fig. 6.8. The minimum among the three techniques once again performed better than the average and the actual position. The highest hit rate was achieved on layout 3, followed by layout 2 where layout 1 performed the worst with more than a 5% drop compared to layout 3. Since there were no obstructions in layout 1 the performance took a steep hit in the other two test datasets. As evidenced by the results, choosing

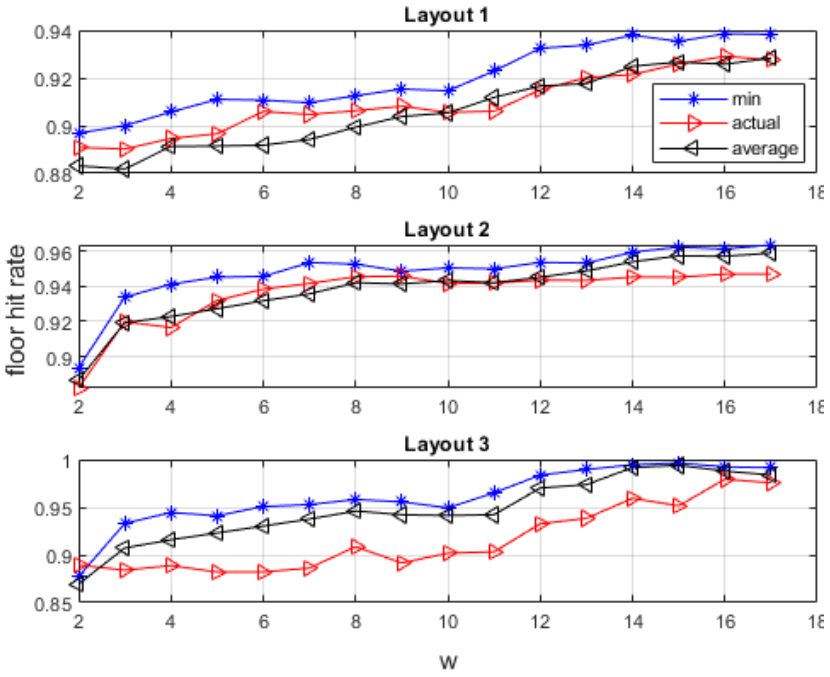
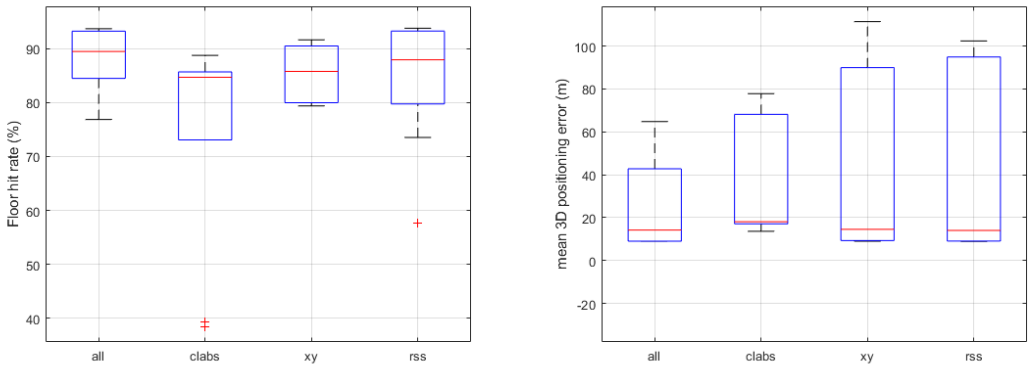


FIGURE 6.9: Floor hit rate for the HDLC dataset.

the best performing transmitter localization technique for each w was the best way to ensure optimal floor and location estimation. Since the dataset was quite small with 17 input features this was one of the few cases where, just the signal strength would have sufficed owing to the increase in features caused by the proposed technique. However, real world systems do not cover a specific corridor and would span multiple APs making the proposed technique viable.

6.3.2 Data normalization and model selection

To test the proposed technique, all possible variations in the data pipeline have to be accounted for to ensure the results are consistent. However, such testing will be too tedious owing to the number of hyperparamters in the pipeline. The variables to be considered here range from the transmitter localization technique, data normalization, model used for training, etc., In order to address this problem the performance of the proposed technique is tested with individual variations maintaining all other variable to be constant. This will account for variation in the parameter alone and help ascertain its influence on the overall performance. The best performing hyperparamters will then be used for testing across other datasets and feature extraction techniques in the subsequent sections.

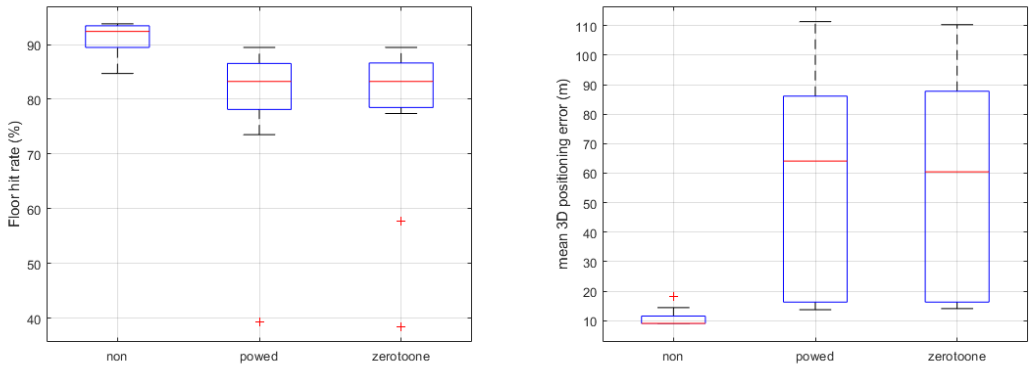


(a) floor hit rate (b) mean position error

FIGURE 6.10: Effect of data choice for UJI dataset

The data choice, which refers to features used to train the model, is an important parameter to identify since there are several combinations owing to the nature of the proposed feature extraction technique, where the number of input features is dependent on the number of APs being considered. The number of APs was chosen to be five for testing and the extreme gradient boosting (xgboost) model was used. The results from testing across all three transmitter localization techniques is shown in Fig. 6.10, where *all* refers to the entire set of input features, rss refers to just the signal strength values of the chosen APs and no structural information, xy and $clabs$ refer to the rss values along with the x and y coordinates for xy and with building and floor features for $clabs$. These groupings were chosen owing to the logical separation of data in each, which can also be used to identify feature importances in each test case. The floor prediction accuracy is highest when all the input features are used for training as shown in Fig. 6.10(a), where the median accuracy is nearly 90%. The other choices also produce similar medians between 85 and 90. This however may not be reflective of the models having learned a universal relationship between the input and output features,owing to a higher spread and outliers in the case of $clabs$ and rss. In the case of xy, the median is lower but the spread is smaller.

When considering the mean three dimensional positioning error shown in Fig. 6.10(b), the trends observed much more pronounced. The worst result produced is significantly better, nearly 20 meters, than the nearest competitor which is $clabs$. The rss grouping suffers a drastic drop in performance compared to floor prediction falling behind all other choices owing to the lack of structural information. We can see that the results are logically consistent with the features and their meanings given the $clabs$ has more connection with three dimensional distances between signals when compared to xy since whether



(a) floor hit rate (b) mean position error

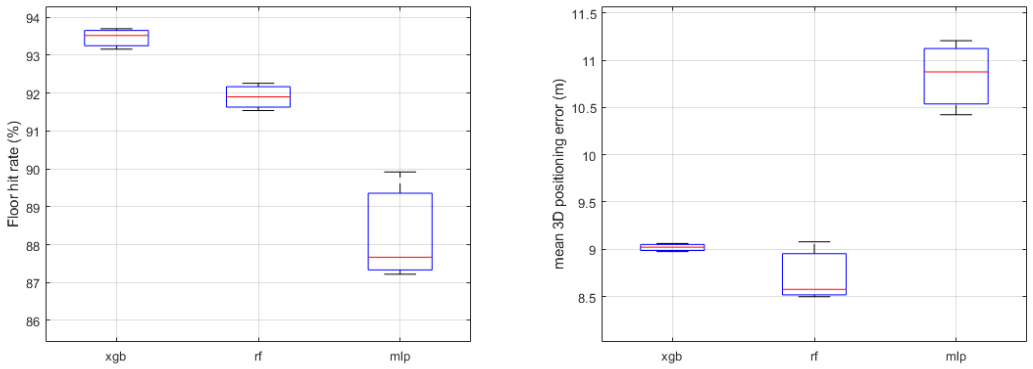
FIGURE 6.11: Effect of normalization technique for UJI dataset

or not APs are on the same floor or building bears more relevance to indoor positioning compared to the raw difference in distances. The fact that the AP locations used are estimated and their accuracy can not be measured, while the estimation accuracy of floors and building will naturally be more accurate than the exact position which is where the x and y coordinates are derived. Thus, using all the input features is clearly the better choice and omitting features leads to a significant drop in accuracy.

The normalization of inputs can help with quicker optimization in some cases and the type of normalization used determines how the data is spread and can be used to emphasize certain differences in the input values. In order to test this, two commonly used normalization techniques, *powed* and *zerotoone*, were compared with the control of *non*, which is non-normalized data as shown in Fig. 6.11. These are chosen since they are commonly used normalization techniques with WiFi signal strength values [68, 137]. The *powed* normalization equation is given as follows.

$$powed = \begin{cases} \left(\frac{RSS_i - min}{-min}\right)^{\beta} & \text{if } i \text{ is from heard AP} \\ 0 & \text{otherwise} \end{cases} \tag{6.9}$$

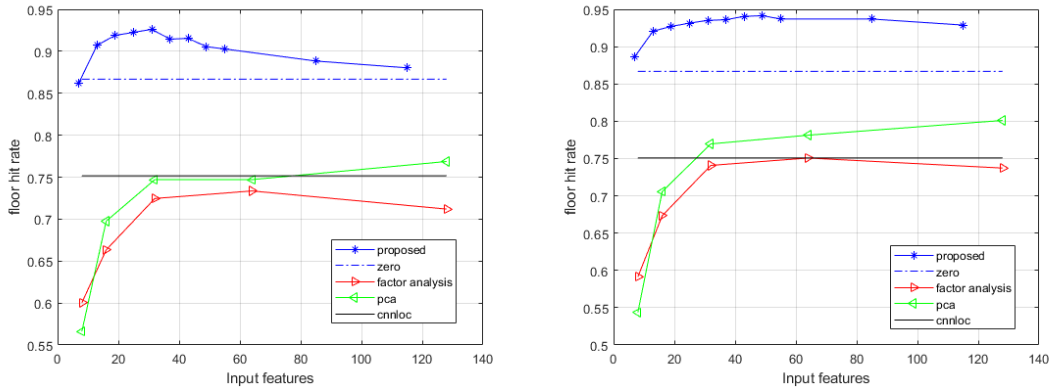
where RSS_i is the signal strength from a list of all heard APs and min is the minimum signal strength recorded. The values for individual input features are scaled to fit the range between zero and one in the case of *zerotoone*. The floor hit rates for the different normalization techniques are shown in Fig. 6.11(a), where the data with no normalization performs the best with the lowest achieved accuracy being more than 80% which was not attained in the other two normalization techniques. The differences are far more



(a) floor hit rate (b) mean position error

FIGURE 6.12: Model testing for UJI dataset

pronounced in the case of mean positioning error as shown in Fig. 6.11(b), where the median is less than 10 meters in the case of no normalization while both the normalization techniques produce nearly 60 meters of error. Though these normalization techniques are commonly used for signal strength data, in this case structural information ranging from distance in meters to difference in floors and buildings are also included which is why the non-normalized data performs better. The model used to learn a relationship between the input features and the normalized output is one of the most important parts of the technique. The results from testing three different models are shown in Fig. 6.12, where xgb is xgboost, rf is random forest and mlp is a multi layer perceptron(MLP). The model parameters are set to the defaults available out of the box in the case of scikit-learn[99] and xgboost[97]. The number of trees for the tree-based techniques were set to 200 and the neural network was trained with three layers of 300 nodes each. This is done to ensure that the models are not limited on their capacity to learn complex relationships owing to the lack of optimization. While optimizing for each dataset is guaranteed to produce better results, the transferability and the ease of implementation are important to enable swift large scale deployment. The floor hit rate is highest for xgboost as shown in Fig. 6.12(a), where the random forest model achieves second highest accuracy of 91.8% while MLP does not manage to breach 90 even in the highest value for that case. The mean 3D positioning error is 8.7 meters, which is the lowest in the case of random forest as shown in Fig. 6.12(b), followed by xgboost which manages 9 meters. MLP produces poorer results compared to the tree-based techniques as observed by Grinsztajn.et.al.[100], while taking 176 seconds for training compared to 77 seconds for random forest and 7.6 seconds for xgboost making it difficult to justify though there are several techniques which manage state-of-the-art performance using such deep learning techniques[68, 116, 123]. The best



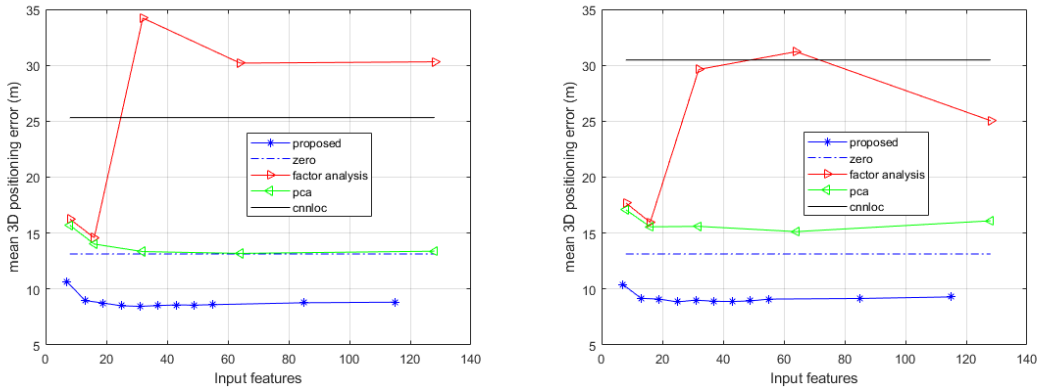
(a) random forest (b) xgboost

FIGURE 6.13: Floor hit rate for UJI dataset

out of the box model for floor estimation is xgboost and the best one for overall position regression is random forest.

6.3.3 Effect of feature extraction

Dimensionality reduction of a sparse RSS vector is the most challenging part of WiFi fingerprinting based indoor positioning. Techniques that perform feature extraction before classification or regression produce better results compared to those techniques that do not use it [29]. Deep learning techniques such as autoencoders or convolutional neural networks (CNN) are capable of feature extraction in the conventional sense of this problem where the input is a sparse vector and the output labels are fixed and specific to the dataset being trained on. However, when the number of input features is fixed and the output is normalized, the problem is reduced to a standard tabular dataset for which tree-based models outstrip deep learning techniques [100]. To study the influence of the proposed feature extraction technique on the positioning performance some commonly used and successful techniques are compared with the proposed technique. Two tree based techniques, random forest and extreme gradient boosting (xgboost), are used to train models to ensure the effect of the model is accounted for in the observed results. The commonly used dimensionality reduction techniques compared with the proposed technique are principal component analysis (PCA) and factor analysis (FA). Apart from these the stacked autoencoder (SAE) trained to reduce the RSS vector to 128 features in the popular CNNLoc paper[68] is also used. However, the CNNLoc paper reports best

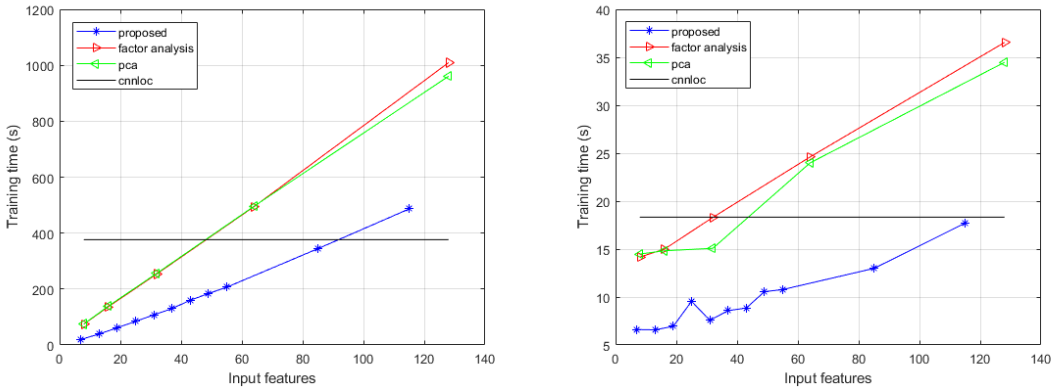


(a) random forest (b) xgboost

FIGURE 6.14: Mean 3D positioning error for UJI dataset

classification accuracy in the case of 128 features, so the value for 128 features is shown in the graphs irrespective of the number of features tested for the other techniques.

The results obtained from testing all the dimensionality reduction techniques and the models on the UJI dataset will be explored in this section. The floor hit rate for random forest is shown in Fig. 6.13(a), where from the legend we can see five variations are observed. PCA outperforms FA and cnnloc at 128 features however is about 10% lower than zero ,which is predicting the strongest AP floor to be the RP location. It is marked zero since this is the value before output de-normalization, this delineates the effectiveness of the proposed feature extraction technique. The proposed technique is shown to improve further on the zero prediction scheme achieving maximum floor accuracy of 92.6% . The zero technique achieving 86.6% indicates that the dataset is imbalanced, which can be addressed to further improve accuracy which is not done here to show out of the box performance without any optimization. Floor hit rate achieved using xgboost is shown in Fig. 6.13(b), where the results are higher than those achieved using random forest for all the dimensionality reduction techniques tested. The maximum floor accuracy with xgboost for the proposed technique is 94.15% while zero is still the closest competitor followed by PCA with 128 features at 80.11% still more than 6% lower than zero prediction. The mean 3D positioning error is used to test overall localization performance with results for random forest shown in Fig. 6.14(a). The proposed technique achieves best results of 8.45 m while the zero prediction scheme results in 13.13 m. The closest of the other dimension reduction techniques is PCA with 13.38 m at 128 features which still performs marginally worse than zero prediction. The results for xgboost are shown in Fig. 6.14(b), where the performance is slightly worse than the random forest counterpart.



(a) random forest (b) xgboost

FIGURE 6.15: Training time for UJI dataset

The FA result improves from 30.31 m to 25.05 m for xgboost which is the only improvement between the models. The PCA performance drops to 15.13 m while the proposed method decreases to 8.84 m. The SAE from cnnloc achieves 25.33 m in conjunction with random forest and 30.47 m with xgboost which is nearly two times worse than zero prediction. While we can see that the same SAE achieved 11.78 m with CNN for regression after parameter optimization, this still is much lower than the proposed technique without any optimization. The reduction of input features reduces the training time, size and complexity of the model. This is important since most WiFi fingerprinting models are deployed on limited hardware such as smart phones and in conjunction with other techniques such as pedestrian dead reckoning[120] or visible light positioning[80]. The time taken to train a model is a good indicator of the computational complexity, which is shown for random forest in 6.15(a). The proposed technique takes 421.8 s with 115 features while cnnloc takes 376.3 s with 128 features, but the PCA and FA fall behind taking more than 960 s to train the model. However, since the best positioning performance with the proposed technique was observed with 31 input features, which takes 84.04 s to train, a fair comparison with the rest shows that the proposed technique still manages to outperform the other techniques. The zero prediction technique which was the second best performing method in previous comparisons does not require a model meaning its training time is also zero. The training times for xgboost are shown in 6.15(b), where we can see a 30 fold drop in training times even for the most time consuming techniques. The proposed technique performs the best across all the techniques here beating out cnnloc marginally with 17.7 s to 18.3 s. The order of the other techniques remains the same as observed for random forest but all being faster. The proposed technique performs the best across all the tested dimensionality reduction techniques with xgboost producing better

floor accuracy at much faster times compared to random forest and random forest producing lowest positioning error. The zero prediction scheme provides a good alternative to the proposed technique coming in second for accuracy and positioning if computational load needs to be as low as possible.

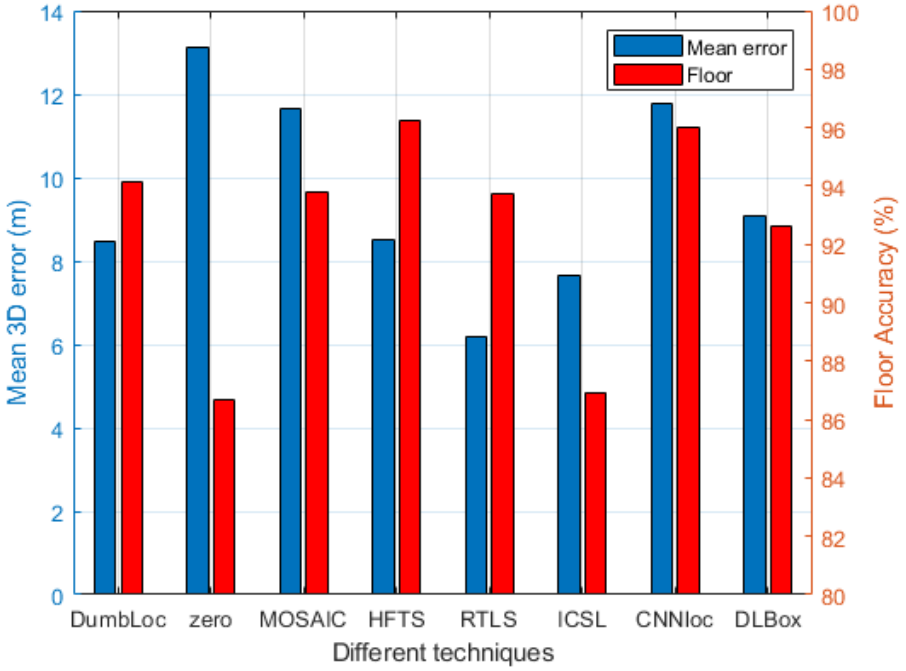


FIGURE 6.16: Comparison of proposed technique with state of the art techniques.

The floor classification accuracy and mean 3D position error from the chosen model was compared to other models trained and tested on the UJI dataset. The proposed technique here is labelled DumbLoc in Fig. 6.16. MOSAIC, HFTS, RTLS and ICSL are winners of the IPIN contest in 2015 [138]. CNNLoc [68] uses a complicated model of a stacked autoencoder followed by 1DCNN and a few dense layers for classification. The DeepLocBox model [139] uses an initial bounding box for area localization. The results of zero prediction are marked zero in the Fig. 6.16. Of these models only CNNLoc and HFTS perform better than the proposed model with both being about 1% better for the floor prediction accuracy. However, both of them produce higher mean error when compared to the proposed technique. The ICSL and RTLS models achieve lower positioning error but produce lower floor classification accuracy. While the zero prediction technique does perform the worst by both metrics across all the models compared, it still manages to remain competitive despite not being trained showing the importance of the feature extraction scheme proposed. The proposed technique was able to achieve these results

with no optimization, using less than 15 APs as opposed to 520 for the others. The zero prediction technique achieves 100% building hit rate for the UJI dataset and since this is the only dataset with multiple buildings, no further training or modelling was required for building estimation. The proposed feature extraction technique was shown to be better than extant popular techniques however, it assumes accurate knowledge of the transmitter locations in the building and a sufficient complexity of the environment in terms of the density of APs in the building. If these AP locations are not available, then the model requires sufficient fingerprinting information to localise the transmitters first and then implement the proposed techniques. While the proposed techniques does not outperform all models in both the mean positioning error and floor accuracy metrics, it is implemented with a much lighter model using far fewer features and a light enough model to be run on limited smartphone hardware. The only restriction imposed by this technique though is the knowledge of transmitter locations to produce such accurate results. In the case of the UJI dataset since the AP locations are not provided, the fingerprinting dataset was used to localise the transmitters first and then perform positioning. The added computation that needs to be performed for this step in the proposed technique is not performed for any of the other state of the art models compared in this section which adds to the complexity of the proposed technique.

6.3.4 Analysis of feature importances

While the proposed technique is shown to perform well on the UJI dataset, ensuring the model has learnt a relationship between the appropriate input features and the normalized output is essential to producing similar results across all such datasets. Most machine learning techniques are black box modelling techniques, which make understanding these relationships even more challenging. The random forest technique however has certain feature importance metrics that can be employed to identify the importance of individual features to the overall performance[96]. These metrics are commonly used to determine feature importance and establish interpretability in random forest models[140]. Random forests are an ensemble of decision trees so the gini metric which is used to construct nodes of a decision tree are used for the mean decrease in impurity(MDI) or gini importance[141]. The MDI feature importance calculates the number of times a feature is used to split a node and weights it by the number of samples it splits. MDI was calculated for the random forest classifier trained for floor estimation on the UJI dataset shown in Fig. 6.17. The data processing here employs 6 APs and the RSS weighted centroid

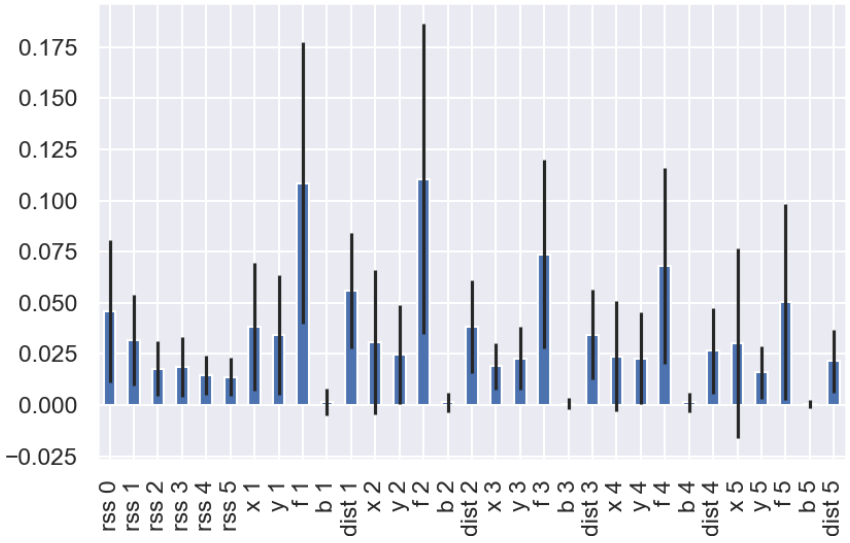
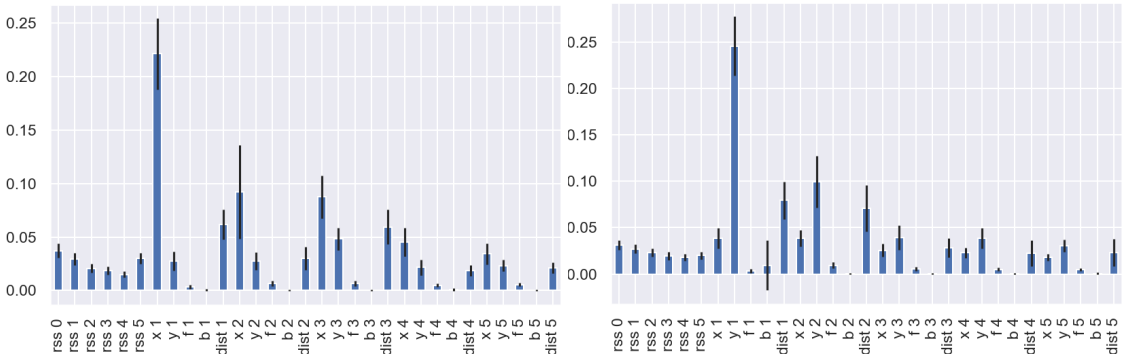


FIGURE 6.17: Mean decrease in impurity of floor estimation for the UJI dataset.



(a) x coordinate (b) y coordinate

FIGURE 6.18: Mean decrease in impurity in position regression for the UJI dataset.

transmitter localization technique. Signal strength of the 6 APs is marked $rss0$ to $rss5$, the difference in x, y, floor, building and euclidean distance between the strongest AP marked $rss0$ and the 5 other APs are also included as the other features. As shown in section 6.3.3, since the building classification accuracy using the zero prediction technique produces 100% accuracy the strongest AP is always in the same building as the RP. The building feature importances, marked $b1$ to $b5$, hold no influence on the output for the floor estimation model. The highest feature importance is observed for the floor features, marked $f1$ to $f5$, in the descending order of importance. The structural feature extracted to indicate the floors holds the highest importance showing the relevance of the data processing proposed here. The distance features, ranging from $dist1$ to $dist5$, also show a similar descending pattern of importances but are not as pronounced as the floor features.

To test the feature importances for regression models employed to estimate location, the MDI for the random forest regressor trained on the same data as earlier was used shown in Fig. 6.18. The MDI of x coordinates is shown in Fig. 6.18(a), where $x1$ is the most important feature with MDI more than twice as much as the second highest feature, which is $x2$. The x coordinate features present similar to the floor features in the floor estimator, but the difference is more pronounced in this case owing to this being a regression problem. The input feature values are not normalized and since the values are in the same scale as the output, the appropriate features are judged to have a much higher importance in this case. Apart from the x coordinate features, the Euclidean distance between the APs is also found to have high feature importance. The MDI of y coordinates is shown in Fig. 6.18(b), where a similar set of feature importances are obtained. In both the coordinates, the feature importances beyond the first three APs have minimal impact on the overall performance. The floor and building values have low feature importances owing to the two dimensional coordinates being regressed. The distance between the first and second strongest APs is found to have a high importance in the prediction of the floor and location, which once again suggests the structural information takes precedence over the signal strength information.

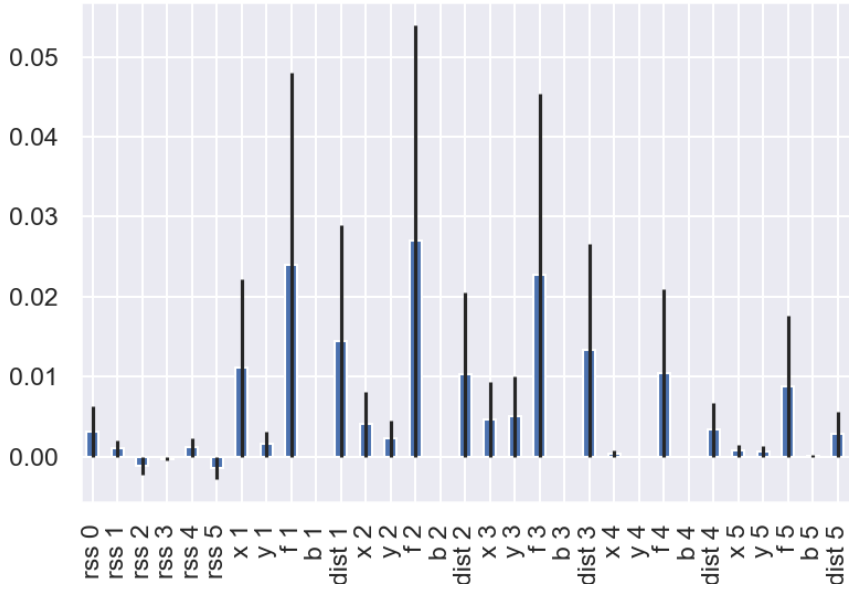
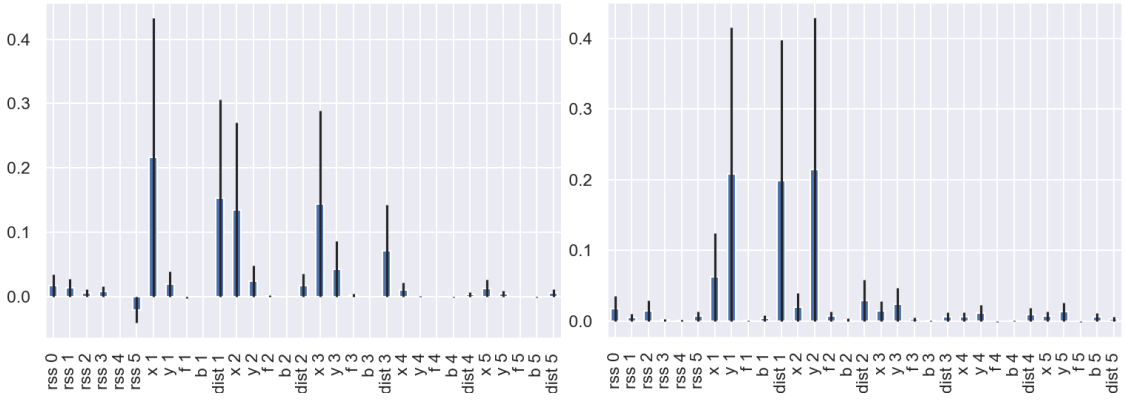


FIGURE 6.19: Permutation importance of floor estimation for the UJI dataset.

Though MDI is useful for feature importance testing, it is susceptible to errors in data with high cardinality[99], this can be avoided by using the other feature importance metric commonly used for random forests, mean decrease in accuracy or permutation importance[142]. This is calculated on the test values by permuting individual features

and averaging the reduction in error caused by the permutation, giving it the name mean decrease in accuracy. Since this is calculated on the smaller test set, concurrence with the expected results can be tested. The permutation feature importance of floor estimation for the UJI dataset is shown in Fig. 6.19, where the most important features are the same as in Fig. 6.17, with the floor difference between the third and strongest AP $f2$ is found to be marginally more important than $f1$. There are some anomalies such as $x1$ which can be expected since the test set is much smaller in the UJI dataset compared to the training set and will not have a wide representation of all possible situations as the training set. However, the most important features are the floor based input features with the distance features following second.



(a) x coordinate (b) y coordinate

FIGURE 6.20: Permutation importance of position regression for the UJI dataset.

The permutation importances of position regression for the UJI dataset are shown in Fig. 6.20, where the x and y coordinates are given separately as in the case of MDI. The decrease in regression accuracy is much more pronounced for both the coordinates than the floor estimation classifier. The x coordinate features and distance features for the first four APs are the only features with a pronounced influence on the x positioning accuracy as shown in Fig. 6.20(a). This is further substantiation of the results for MDI albeit with drastic differences owing to the smaller test set. The y coordinate features and distance features of only the first three APs see high importances in the y positioning model as shown in Fig. 6.20(b). For the test case used here, a lower number of APs could have been used for indoor positioning since the feature importances drop off past the fourth AP as opposed to floor estimation where all the six APs exert influence on the classification accuracy. The signal strength features barely record any variation highlighting the importance of structural features.

TABLE 6.1: Database information

Dataset	Train	Test	APs	# b	# f	Citations	Reference
LIB 1	576	3120	174	1	2	95	[124]
LIB 2	576	3120	197	1	2	95	[124]
SAH 1	9291	156	775	1	3	5	[143]
TIE 1	10633	50	613	1	6	5	[143]
TUT 1	1476	490	309	1	4	72	[144]
TUT 2	584	176	354	1	3	72	[144]
TUT 3	697	3951	992	1	5	128	[75]
TUT 4	3951	697	992	1	5	128	[75]
TUT 5	446	982	489	1	3	15	[145]
UJI	19861	1111	520	3	5, 4, 4	405	[74]
UTS 1	9108	388	589	1	16	125	[68]

6.3.5 Performance generalization testing

Now the 11 datasets being considered are shown in table 6.1, where the train and test refer to the training and test samples in the datasets, # b and # f represent the number of buildings and floors in each dataset. The citations column outlines the number of papers each of these datasets has been cited in. The UJI dataset is the only dataset with multiple building data, also has the most number of training samples and it has been cited far more than any other with 405 citations compared to 128 for the next best. Hence the initial training process outlined as the offline process in Fig. 6.1, is performed using the UJI dataset and the subsequent online process will be performed using all the datasets. Note that all of these models have several APs ranging from 174 to 992 and that we will only be using the strongest w APs at each RP for all of them.

$$\#IF = w + (5(w-1)) \tag{6.10}$$

where, $\#IF$ is the number of input features and w is the number of APs. The number of input features increases as the number of APs increases but given that the majority of RPs have less than 20 APs being heard[74], testing was one till the same. The varying number of input features must first be normalized in order to use the same model.

$$N\#IF = \frac{1}{N_d}\sum_{i=1}^{N_d}\frac{N_{ap}(i)}{\#IF} \tag{6.11}$$

where, $N\#IF$ is the normalized number of input features, N_d is the number of datasets, N_{ap} is the number of APs in the dataset since this would be the number of input features normally and $\#IF$ is the number of input features which changes as w is varied.

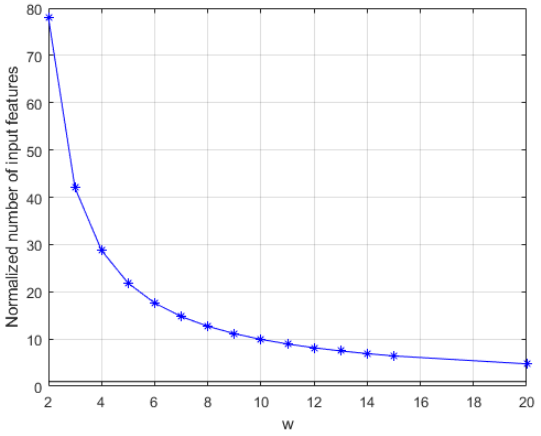


FIGURE 6.21: Normalized number of input features for all the datasets.

The normalized number of input features is plotted as a function of the number of APs chosen in Fig. 6.21. The normalized number of input features across the 11 datasets considered is more than 4 times greater than the proposed technique even when 20 strongest APs are considered for each RP and more than 10 time greater when 10 APs are considered showing the drastic reduction in the number of input features achieved by the proposed feature extraction technique. Since this fraction is calculated across all the datasets and averaged, it gives a robust measure of the multiplier to be applied when all the data is used as would normally be the case.

$$NFH = \frac{1}{N_d} \sum_{i=1}^{N_d} \frac{FH(i)}{FH_{zero}(i)} \tag{6.12}$$

where, NFH is the normalized floor hit rate, FH is the floor hit rate being normalized and FH_{zero} is the floor hit rate achieved using zero prediction technique. This will show how a technique fares compared to the zero prediction technique. The normalized floor hit rate across all the datasets is shown in Fig. 6.22, where the legend indicates 6 different techniques being compared. The floor hit rate corresponding to the random forest and xgboost models are marked rf and xgb respectively. The results from these two models are consistent with the results observed in section 6.3.3, with xgboost producing higher floor prediction accuracy than all other models. The results obtained from using the model pre-trained using the UJI dataset to predict floors for all the datasets is labelled

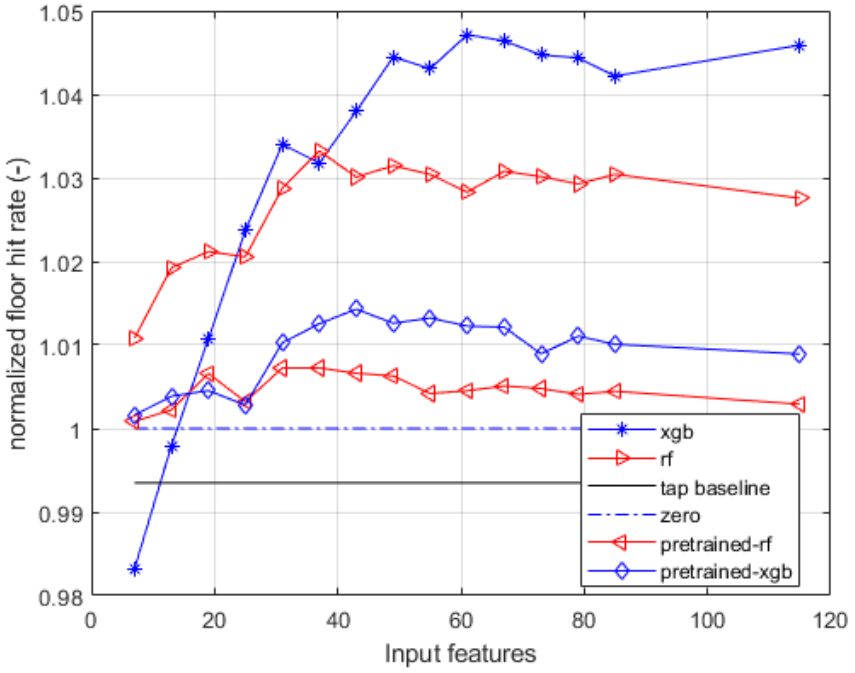


FIGURE 6.22: Normalized floor hit rate for all the datasets.

pretrained-rf or xgb depending on the technique used to train the original model. These are also similar to the prior models trained using the same techniques with xgboost producing better results compared to random forest but they are worse than the corresponding model trained individually on each dataset. These pre-trained models however are better than the zero prediction technique. The floor hit rate obtained by the Klus et.al. [146] after normalization was marked tap baseline on the plot. These are the closest floor hit rate results tested on public datasets and after normalization the results are found to be marginally lower than the zero prediction technique. The tap baseline technique was developed with a focus on the prediction times which are greatly improved by the zero prediction technique not needing to be trained.

The floor hit rate of all the datasets individually is shown in Fig. 6.23, where the tap baseline, xgboost, zero prediction and pre-trained xgboost models are considered among the models tested. The xgboost model was shown to outperform random forest across all the datasets in Fig. 6.22 hence, this technique along with the pre-trained xgboost model from the UJI dataset were used for comparison. Though the normalized floor hit rate provides a representation of the overall performance, the individual datasets and their performance need to be studied to gain a better understanding of the model. The tap baseline performs better than xgboost in seven of the eleven datasets, four of these are

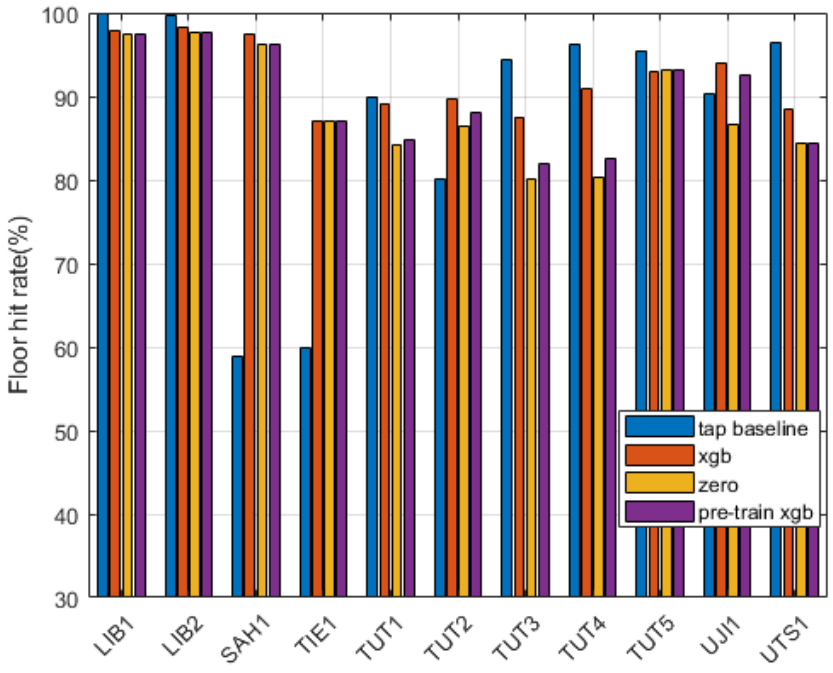
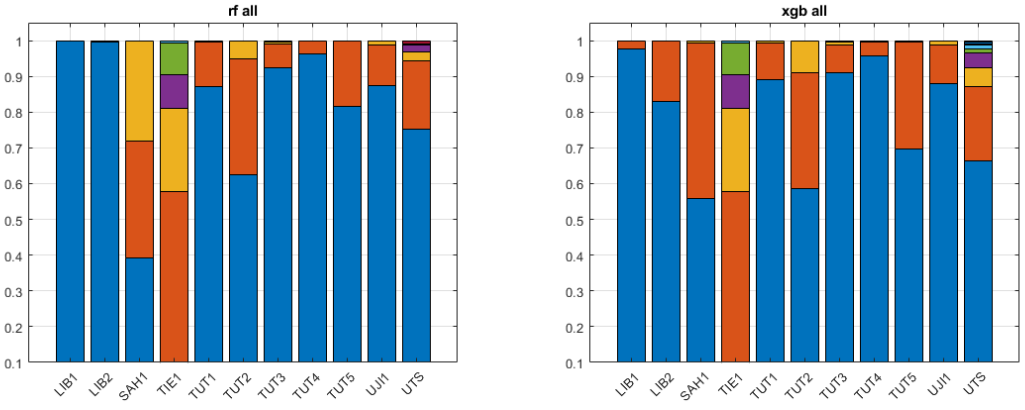


FIGURE 6.23: Floor hit rate for all the datasets.

about two percent better than xgboost. The most pronounced difference in performance is in the crowd sourced datasets, TUT3 and TUT4, and in the UTS dataset with sixteen floors.

Even in these three datasets the differences range from five to eight percent, with UTS producing the biggest difference. This difference can be attributed to the large number of floors in this building, owing to the output classes are limited to four in either direction. While it may be tempting to increase the number of classes, the proposed technique outstrips the baseline by a large margin in the four datasets where former performs better. The smallest difference is in the UJI dataset with a 4 percent difference, followed by the TUT2 dataset with a ten percent difference. The largest improvement compared to the baseline was achieved in the case of SAH1 and TIE1 datasets, with an approximate forty and thirty percent improvement improvement respectively. These datasets are the reason why zero prediction technique achieves higher normalized performance than the baseline in Fig. 6.22. The zero prediction and pre-trained xgboost models achieve higher floor hit rate compared to the tap baseline in only TUT2 apart from these two datasets, but the massive difference in these two datasets puts them ahead of the tap baseline since the baseline does not achieve more than a ten percent improvement in any of the datasets.

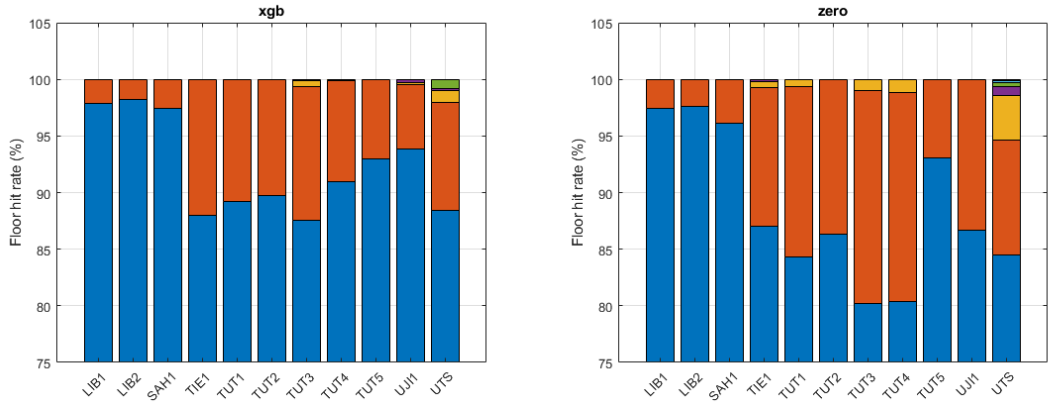


(a) random forest (b) xgboost

FIGURE 6.24: Floor difference without data processing

The zero prediction and pre-trained models show that they are low-cost, low-effort alternatives that perform evenly across all datasets owing to the relationship learnt between the structural information and the relative labels.

While the number of instances predicting the floor accurately is reflected by the floor hit rate, the magnitude of inaccuracy in incorrect prediction is not studied. Though the overall positioning error is studied, most indoor positioning techniques use WiFi in tandem with other techniques and WiFi helps in most cases to estimate the coarse location while other techniques provide fine adjustment[27]. Therefore, understanding how the proposed technique fails in floor estimation will help improve the overall positioning accuracy of the indoor positioning system by reducing reliance on the results. Since this information is not provided in the paper from which the tap baseline was taken, the default models without data processing were tested. The results for model trained with raw data are shown in Fig. 6.24, where the blue bar denotes accurate prediction with stacked bars in order of increasing difference. The red bar denotes a difference of one between the predicted floor and ground truth ,similarly yellow denotes two, purple denotes three and green denotes four. The other colours occur only in the case of UTS since it has sixteen floors. The performance of random forest, shown in Fig. 6.24(a), in this case is better than xgboost for LIB1 and 2, but the performance of both models is much worse than the baseline for the TIE1 dataset with less than ten percent accuracy. This dataset has an extremely small test dataset of fifty samples so the overall number of inaccurate predictions to cause the hit rate to drop is much lower as well, this makes the results achieved by the proposed technique even better. In the SAH1 dataset, random forest suffers a low overall hit rate and the difference of tow floors is observed in more than twenty percent of the



(a) xgboost (b) zero prediction

FIGURE 6.25: Floor difference with data processing

values as compared to no two floor errors in xgboost, shown in Fig. 6.24(b). The other datasets predominantly exhibit one floor error in the case of incorrect estimation. The effect of the feature extraction is highlighted by using the models now with the proposed data processing techniques in Fig. 6.25, where the results are not below 80% for any of the models. The xgboost model results are chosen over random forest owing to the results in Fig. 6.22. The xgboost model results, shown in Fig. 6.25(a), show nearly 99% one floor error hit rate in all the datasets. The results from the zero prediction technique, shown in Fig. 6.25(b), where no machine learning technique was employed achieves the lowest one floor error hit rate of 95% on the UTS dataset, with results reaching 99% on the lower end for all the other datasets. This further solidifies the results in the Fig. 6.23, showing that even when the prediction is incorrect it is seldom by more than one floor.

The mean 3D positioning error was normalized like the floor hit rate with the mean error from the zero prediction technique and shown as a function of the number of input features in Fig. 6.26. The labels in the legend are the same as Fig. 6.22, with the only difference being the knn baseline which are positioning results obtained using k nearest neighbours(knn) technique which was used to compare with the paper[146] from which the floor prediction results were obtained. Since the results achieved by their technique in that paper were marginally lower than the baseline, the proposed techniques are compared with the knn baseline. The random forest technique produces best results closely followed by xgboost, both of which outperform the knn baseline. The positioning error is different from the floor hit rate in that the zero prediction technique on average does not perform better than the baseline. The models pre-trained on the UJI dataset follow the same pattern as observed in section 6.3.3 with random forest model performing slightly

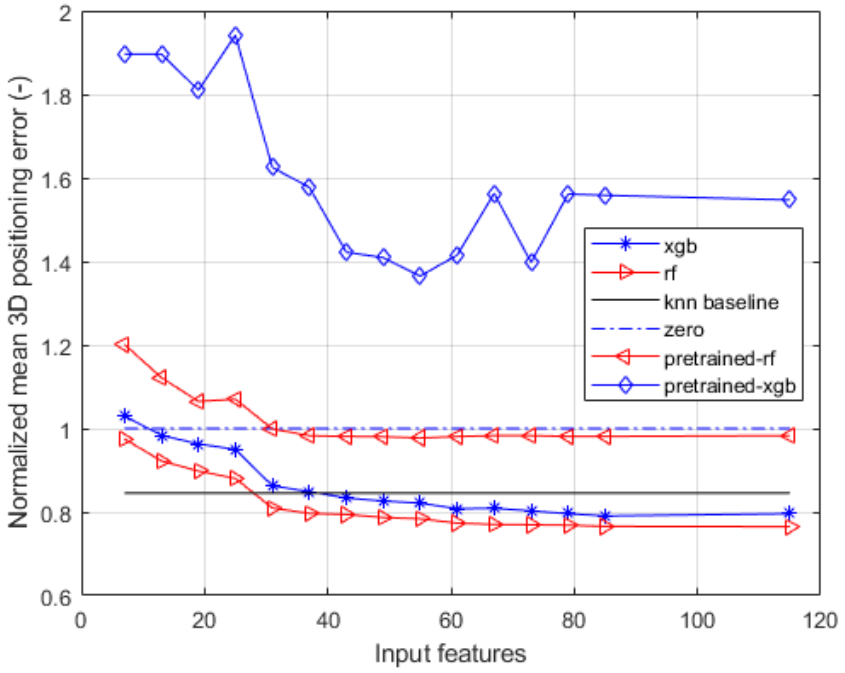


FIGURE 6.26: Normalized mean 3D positioning error for all the datasets.

better than the zero prediction technique but falling short of the knn baseline. The xgboost pre-trained model fares the worst of all the methods tested with results 50% worse than the zero prediction technique.

Similar to the floor estimation using proposed techniques, the mean positioning error in the individual datasets are shown in Fig. 6.27, where results from the random forest model and zero prediction technique were included. The pre-trained models were not included owing to higher normalized error than the baseline and only marginally better performance than the zero prediction technique. There are no major discrepancies in the performance among the models in individual datasets as was observed in Fig. 6.23. The datasets which cover lower area, LIB1 and 2, achieve the lowest error of around 3 meters in all the models and the baseline. Unlike the floor hit rate, the baseline here is better than the random forest results in five datasets and the datasets which contributed to the major difference in floor hit rate, SAH1 and TIE1, are among these five datasets. This highlights the major impact small test datasets can have on accuracy metrics. Since the overall distance is highlighted by the positioning error, the performance remains the same with TIE1 showing a meter improvement for the baseline compared to the random forest model. The crowd-sourced datasets perform better in the overall positioning error as they did with the floor hit rate. This is expected since the AP locations which form the corner

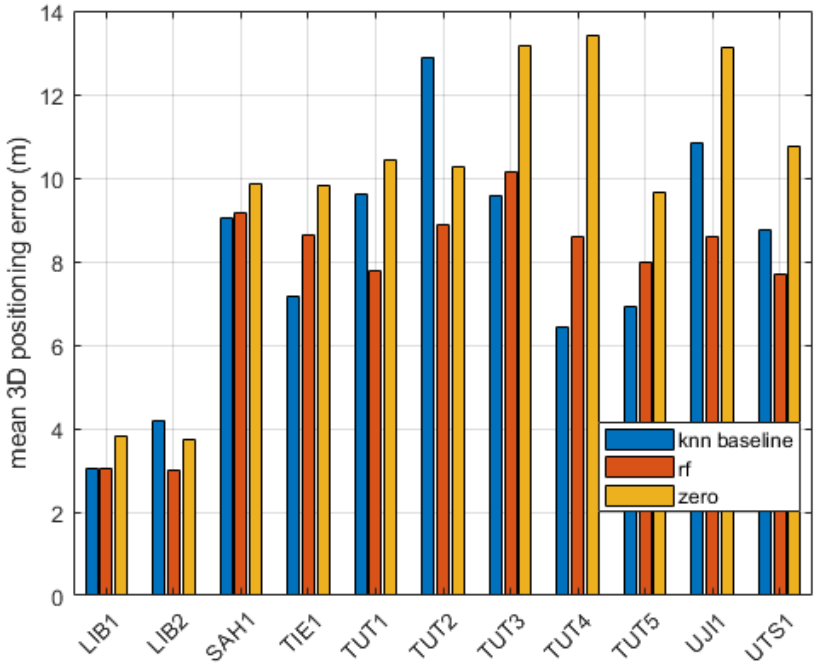
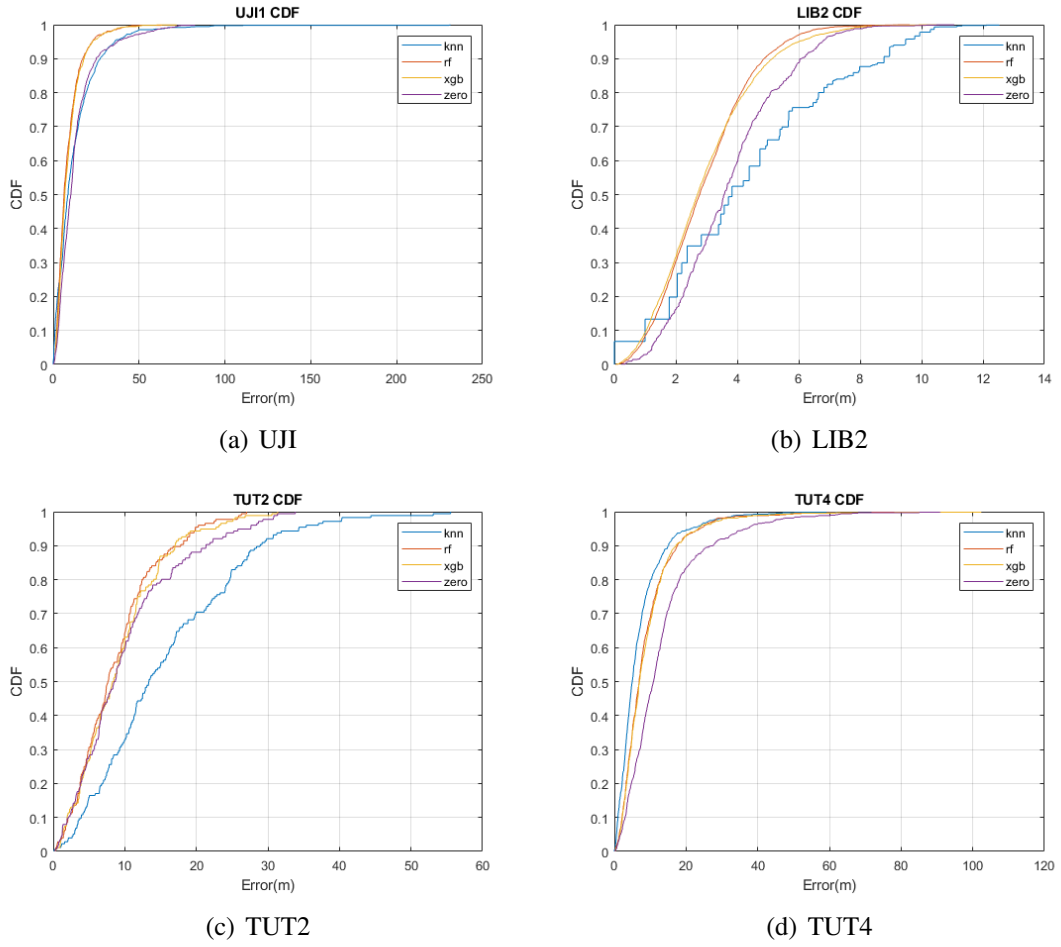


FIGURE 6.27: Mean 3D positioning error for all the datasets.

stone of the proposed technique, providing all the structural information, are bound to be erroneous since the RPs are not recorded accurately. The UTS dataset which covers a large area produces lower error with the proposed technique compared to the baseline as does the UJI dataset. The performance of the zero prediction technique is limited by the coordinate normalization ensuring all the datasets use the same coordinate system. While the technique fails to outperform the baseline in this case, it does achieve results better than the baseline in some cases, TUT2 and LIB2 datasets, highlighting the efficacy of the proposed feature extraction technique. While random forest performs best on large areas, the zero prediction technique is limited in this regard. Though the maximum error is limited by the normalization of coordinate points, the average does increase owing to the strongest AP location being predicted in all cases.

The position error distribution for some individual datasets are shown in Fig. refcdf, where the knn baseline is compared with random forest, xgboost and the zero prediction technique. The datasets were chosen based on the different categories of datasets to highlight the difference in performance among the techniques in these datasets. The UJI dataset is chosen to represent the performance among datasets covering large areas shown in Fig. 6.28(a), where the baseline performs better than the zero prediction technique but worse than both the random forest and xgboost models. The maximum error



(a) UJI (b) LIB2

(c) TUT2 (d) TUT4

FIGURE 6.28: Positioning error distribution in individual datasets

is limited to about fifty meters in the case of these models, while the baseline and zero prediction technique produce errors greater than sixty meters in a small percentage of points. The baseline has a lower mean error than the zero prediction technique since it has a higher probability of lower error. The LIB2 dataset was chosen to highlight the positioning error in datasets covering small areas shown in Fig. 6.28(b), where the baseline performs worse than the zero prediction technique. The error scale is much smaller compared to the UJI dataset highlighting the difference in scale between the datasets. The probability of error being less than three meters is more for the baseline than the zero prediction technique however, the maximum error for the baseline is much higher at around eleven meters compared to nearly eight meters achieved by the zero prediction technique. The random forest and xgboost models outperform the baseline by a significant margin which is reflected in the overall lower mean positioning error. The results of the crowd-sourced datasets were represented using TUT4 in Fig. 6.28(d), where the baseline achieves error lower than twenty meters for about ninty five percent of the points

tested while the zero prediction technique has errors of around forty meters for the same ninty five percent mark. This translates to overall performance where the baseline outperforms random forest by more than a meter managing half the mean error of the zero prediction technique. The scale here is much larger than the LIB2 dataset, which causes such pronounced differences which may not be apparent from the error distribution. The error distribution achieved by a non-crowd sourced dataset is shown in Fig. 6.28(c), where the baseline performs much worse than even the zero prediction technique. The proposed techniques produce similar results for about sixty percent of the points with error less than ten meters, this is however changed at the ninty five percent mark with the random forest and xgboost models showing less than twenty meters error with zero prediction technique only achieving less than thirty meters. The baseline is not far behind at approximately thirty five meters, but since the maximum error stretches to more than fifty meters the mean error is much higher than the other techniques. Thus, we can see that the performance is consistent, when broken down into error distributions within individual datasets.

The proposed technique manages to generalize well to 10 other datasets, producing much better floor hit rate compared the closest competitor without any training in the case of the pre-trained model and the zero prediction technique. The mean positioning results however indicate that the location estimation is not as easy to improve on the baseline. The proposed technique performs better on large datasets with large number of APs as evidenced by the UJI dataset as opposed to the HDLC dataset which is a smaller dataset with only 17 APs. Among the datasets tested TUT 3 and TUT 4 are crowd-sourced, which shows that even if the RP location is not accurately tracked, the proposed technique manages to achieve parity with the baseline for floor level and location estimation. The performance thus shows that the proposed technique enables a building owner with actual AP locations to implement a WiFi fingerprinting based indoor positioning system without any new data, using a model pre-trained on publicly available datasets or using the zero prediction technique. Once the system is deployed more data can be crowd-sourced and labelled using these techniques to implement the proposed technique and optimized if better positioning accuracy is needed.

6.4 Conclusion

This chapter proposes a machine learning-based technique for floor and building estimation without the need for optimization. A novel input and output structure enabled similar (within 2%) floor prediction accuracy and mean positioning error to state-of-the-art models on the UJI dataset using less than 5% of the inputs. The pre-trained model from this dataset and the zero prediction technique, with no machine learning model, were tested on 10 other publicly available datasets and achieved better floor hit rate averaged over the datasets compared to the closest competitor. The proposed technique with fingerprinting data was shown to outperform the current closest competitor in overall positioning accuracy after training. Thus the model learned a relationship between the inputs, strongest AP RSS with their relative locations, and the outputs, normalized floor labels. Since this relationship is universally similar, no new data or training was needed to achieve high floor prediction accuracy on completely new datasets. The results are testament to the superior generalization achieved by this technique.

These tests were conducted to explore the viability of WiFi as a floor estimation and coarse location prediction scheme. Visible light positioning(VLP) systems relying on CMOS sensors do not scale well to large areas owing to their limitations in visible light communication(VLC), which is essential for coarse localization. While VLP produces highly accurate location estimates, in the order of centimeters, with respect to a transmitter, VLC is needed to identify the location of the light. Commercial off the shelf(COTS) lights are not capable of the MHz switching rates which are needed to establish VLC and the lights that are capable of the same are expensive. Even if the expensive lights were used sparingly, the CMOS sensors on smartphones will not be capable of reading them. While simpler coding and modulation schemes can be employed at lower frequency ranges, they become expensive to deploy and maintain on a large scale. Thus, employing WiFi as an input modality solves the coarse localization problem without any new hardware deployment since the technique uses existing APs used for communication. Since the proposed technique focuses on generalization and transferability, the coarse localization accuracy, floor and building localization, achieved is consistently high across the different datasets tested. The zero prediction technique is particularly useful in the case of fusion with other VLP techniques on a smartphone, which involves computationally intensive image processing. In the case of AP locations being available, accurate floor and building localization can be achieved without any fingerprinting or data collection.

Chapter 7

Conclusion and future work

7.1 Conclusions

We explored indoor visible light positioning techniques using smartphone cameras and sensors with a focus on leveraging cheap and extant infrastructure to facilitate robust and accurate indoor positioning.

We explored a 3D geometry based computer vision algorithm for indoor positioning in Chapter 3. We used common properties of rectangular panel lights and tested the mean positioning error at four heights ranging from 1.2 to 1.56m from the transmitter in 12cm increments. The mean 3D positioning error was found to be better than the current SOTA solution producing 0.74, 0.85, 0.92, 1.12 cm errors at 1.2, 1.32, 1.44 and 1.56m from the transmitter respectively. The angular errors were also analysed by randomly varying the orientation of the smartphone when capturing images at all four heights. The IMU on the smartphone was used to determine the heading and hence estimate the orientation from extracted features. The orientation errors were also found to be lower for the proposed technique compared to the SOTA technique producing 3.32° azimuth error compared to 7.86° for the SOTA. This technique was further improved to address the partially visible lights in an image by estimating corners based on the portion of the light visible in the image. This was the first technique to perform positioning on partially visible lights where only two of the four corners were visible in rectangular panel lights. The proposed technique achieved 2.27 cm mean 3D positioning error and 3.57° azimuth error.

In Chapter 4, we proposed a machine learning based indoor positioning technique using simulated images. We used Blender to simulate photo-realistic images which were then used to train machine learning models to estimate indoor position. The models were trained on corners from simulated and real images to compare results. The models were tested on data gathered at four different heights from the transmitter. The tree based models were shown to outperform traditional neural networks for this problem. The model trained on simulated data produced a mean 3D positioning error of 3.99 cm while the model trained on real images produces 0.72 cm. Though the model trained on simulated data outperformed the SOTA technique with 6.68 cm, it still performed worse than the computer vision technique and the model trained on real data.

Both the techniques till now have explored positioning with respect to the transmitter position, this assumes the position of the transmitter is known which enables sub cm positioning accuracy in some cases. However a real world implementation of these techniques will require a technique to estimate the transmitter location. We explored the feasibility of transmitter localisation using optical camera communication in Chapter 5. We proposed a technique to simulate OCC on transmitters and compared it with a SOTA simulation technique. The discrete Fréchet distance values for the two simulation techniques with experimental images were used to show that the proposed simulation technique was better than the SOTA technique. The worst DFD values for the proposed technique were less than half of the best DFD values for the SOTA technique. The simulated images and experimental images were decoded using a conventional demodulation technique and the resultant detection success rate was found to be similar across different transmitter and receiver properties. These were tested with varying exposure times, switching frequencies from two different cameras on the receiver side and varying shape, size and luminous intensity through two different transmitters. BER was analysed from a one minute video at difference distances from the transmitter for two different modulation and encoding techniques. The BER was shown to be lower than the hard decode level needed before forward error correction using the proposed demodulation technique.

While OCC is a viable technique for transmitter localisation, it still needs additional electronics to be installed to switch the lights on and off at high frequencies to implement simple modulation schemes. Even the cheap simple electronics become expensive to install and maintain when scaled to thousands of lights as encountered over a mall or university campus. Hence, we used Wi-Fi fingerprinting to estimate coarse location which can used for transmitter localisation. Since fingerprinting is a time and resource

intensive process, we leveraged existing open source datasets to train models for building, floor and position estimation. A novel feature extraction technique was used to reduce the amount of features required for training without compromising on positioning accuracy. The proposed technique was shown to be better than other feature extraction techniques. The extracted features were then used to train tree-based machine learning models of which xgboost produced the highest floor accuracy and random forest produced the best positioning accuracy. On the benchmark UJI dataset, the proposed technique produced 94.15% floor prediction accuracy and 8.45m mean 3D positioning error using less than 5% of the AP values. The proposed technique was tested on eleven datasets to show the generalisability of the technique. This technique was reused on ten other public datasets and achieved an average floor estimation accuracy of 91.93% when trained with new data and 88.68% without any new data compared to 87.1% of the closest competitor.

7.2 Future Work

While the camera based indoor visible light positioning techniques outlined in this thesis deal with existing infrastructure and minimizing deployment and maintenance costs, the four techniques are presented by themselves without studying the effects of implementing them together. These four techniques are split into coarse positioning and fine positioning where both are required to implement a scalable indoor positioning system. The computer vision based positioning technique improves on current techniques by enabling positioning for rectangular panel lights when only two corners of the light are visible. However, the extent to which the light has to be visible in the image and the influence of the partial visibility on the positioning accuracy can be further explored. The easy to use simulation technique using Blender provides a way to generate a lot of images which will be needed to train deep learning models for position regression. We used a simpler technique where image processing is used to simplify the problem before using machine learning models owing to limited computational resources available at our disposal. The use of this simulation technique to render images for direct viewpoint estimation using deep learning would be a natural extension of the presented work.

For the coarse positioning techniques, possible solutions are suggested but the actual deployment and testing of these techniques in tandem with the fine positioning techniques has yet to be explored. This is a possible avenue for expansion of the work presented

in this thesis. The OCC chapter covers a simulation technique to generate data for testing decoding algorithms, where massive amounts of images that can be generate open the door for deep learning based direct demodulation and decoding from images. The work presented uses image processing to simplify the problem before employing machine learning due to limited computational resources at our disposal. When direct demodulation in the coarse localisation and direct regression for fine localisation are possible they can be fused together to generate one end to end deep learning model capable of accurate and robust positioning. New modulation techniques covered in the amendment to IEEE 802.15.7-2018 standard by the IEEE 802.15.7a (TG7a) task group greatly improve the data rate and mobility [147]. The physical layer definition PHY-7 was shown to achieve several Mb/s bridging the gap between VLC and OCC and mobility nearing 10 km/h which was not possible for any of the previously suggested schemes. Since we could not afford the equipment required to implement and test these schemes, these complicated schemes were not tested. Further research on such techniques will hasten the deployment of OCC for indoor positioning.

Though OCC was explored in detail, the photodetector based schemes are closer to commercial adoption with the IEEE 802.11bb standard. The light antenna ONE from pureLiFi has been made available to original equipment manufacturers for testing. Once this makes its way on to commercial devices such as smartphones, the interaction of both photodetectors and cameras will need to be studied for robust and accurate indoor positioning systems. This further opens the door for studying the simultaneous use of visible light communication through photodetectors which allow for fast data rates while cameras allow for accurate visible light positioning owing to the directional information available from cameras. The influence of functions such as handover on quality of service metrics can also be further explored. The proximal policy optimisation for hybrid LiFi and WiFi indoor networks can be improved using the proposed simulation technique for testing. The proposed simulation technique can be used to test the coverage provided for different transmitter and receiver properties allowing for optimal modulation, transmitter procurement and positioning. Current light panel positioning optimisation schemes explore the effect of illumination, while the proposed simulation scheme can be used to choose and place the light panels with the indoor positioning and communication needs.

The proposed WiFi positioning scheme improved on the current state of the art for building, floor and position estimation. This was proposed to aid quick implementation without the need for painstaking fingerprinting covering large areas. However, in cases where

the transmitter location is not available or not provided due to security concerns a drop in positioning accuracy will be unavoidable. The initial model can be improved after deployment using user data through federated learning[148, 149]. Further research into the integration of federated learning with the proposed positioning technique would be beneficial for both WiFi based positioning and VLP. With both the coarse and fine positioning techniques implemented, their integration through sensor fusion using other sensors commonly available on smartphones can greatly improve the robustness, scalability, speed and accuracy of indoor camera based VLP systems.

List of Author's Awards, Patents, and Publications[1]

Journal Articles

- **S.C.Narasimman** and A.Alphones, "DumbLoc: Dumb Indoor Localization Framework using WiFi Fingerprinting," *IEEE Sensors Journal*, 14623-14630, vol 24, issue 9, (2024).
- **S.C.Narasimman** and A.Alphones, "Indoor Visible Light Positioning for A Single Partially Visible LED," *IEEE Sensors Letters*, 1-4, vol 8, issue 5,(2024).

Conference Proceedings

- **S.C.Narasimman** and A.Alphones, "Tree-based Single LED Indoor Visible Light Positioning Technique," in *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*, 826-831(2023).
- **S.C. Narasimman**, and A. Alphones. "Simulation and Experimental Validation of Optical Camera Communication," in *TENCON 2024 - 2024 IEEE Region 10 Conference (TENCON)*, 1135-1138(2024).

[1]The superscript * indicates joint first authors

Bibliography

[1] Hui Liu, Houshang Darabi, Pat Banerjee, and Jing Liu. Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man and Cybernetics, Part C (Applications and Reviews)*, 37(6):1067–1080, nov 2007. doi: 10.1109/tsmcc.2007.905750. 1, 2, 76

[2] Chouchang Yang and Huai-rong Shao. Wifi-based indoor positioning. *IEEE Communications Magazine*, 53(3):150–157, March 2015. ISSN 0163-6804. doi: 10.1109/mcom.2015.7060497. 1

[3] Rui Ma, Qiang Guo, Changzhen Hu, and Jingfeng Xue. An improved wifi indoor positioning algorithm by weighted fusion. *Sensors*, 15(9):21824–21843, August 2015. ISSN 1424-8220. doi: 10.3390/s150921824.

[4] Nicolas Le Dortz, Florian Gain, and Per Zetterberg. Wifi fingerprint indoor positioning system using probability distribution comparison. In *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, March 2012. doi: 10.1109/icassp.2012.6288374.

[5] Ye Tao and Long Zhao. A novel system for wifi radio map automatic adaptation and indoor positioning. *IEEE Transactions on Vehicular Technology*, 67 (11):10683–10692, November 2018. ISSN 1939-9359. doi: 10.1109/tvt.2018.2867065. 1

[6] Samer S. Saab and Zahi S. Nakad. A standalone rfid indoor positioning system using passive tags. *IEEE Transactions on Industrial Electronics*, 58(5):1961–1970, May 2011. ISSN 1557-9948. doi: 10.1109/tie.2010.2055774. 1

[7] He Xu, Ye Ding, Peng Li, Ruchuan Wang, and Yizhu Li. An RFID indoor positioning algorithm based on bayesian probability and k-nearest neighbor. *Sensors*, 17(8):1806, aug 2017. doi: 10.3390/s17081806. 1, 76

[8] Lu Bai, Fabio Ciravegna, Raymond Bond, and Maurice Mulvenna. A low cost indoor positioning system using bluetooth low energy. *IEEE Access*, 8:136858–136871, 2020. doi: 10.1109/access.2020.3012342. 1, 76

[9] C. K. M. Lee, C. M. Ip, Taezoon Park, and S.Y. Chung. A bluetooth location-based indoor positioning system for asset tracking in warehouse. In *2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*. IEEE, December 2019. doi: 10.1109/ieem44572.2019.8978639.

[10] Naga Sai Ravali Challa, Padmapriya Kesari, Supraja Reddy Ammana, Satyanarayana Katukojwala, and Dattatreya Sarma Achanta. Design and implementation of bluetooth-beacon based indoor positioning system. In *2019 IEEE International WIE Conference on Electrical and Computer Engineering (WIECON-ECE)*. IEEE, November 2019. doi: 10.1109/wiecon-ece48653.2019.9019997.

[11] Cheng Zhou, Jiazheng Yuan, Hongzhe Liu, and Jing Qiu. Bluetooth indoor positioning based on rssi and kalman filter. *Wireless Personal Communications*, 96 (3):4115–4130, July 2017. ISSN 1572-834X. doi: 10.1007/s11277-017-4371-4. 1

[12] Sheng-Cheng Yeh, Wang-Hsin Hsu, Wen-Yen Lin, and Yi-Fan Wu. Study on an indoor positioning system using earth's magnetic field. *IEEE Transactions on Instrumentation and Measurement*, 69(3):865–872, March 2020. ISSN 1557-9662. doi: 10.1109/tim.2019.2905750. 1

[13] Seong-Eun Kim, Yong Kim, Jihyun Yoon, and Eung Sun Kim. Indoor positioning system using geomagnetic anomalies for smartphones. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, November 2012. doi: 10.1109/ipin.2012.6418947. 1

[14] Thomas Gigl, Gerard J.M. Janssen, Vedran Dizdarevic, Klaus Witrisal, and Zoubir Irahhauten. Analysis of a uwb indoor positioning system based on received signal strength. In *2007 4th Workshop on Positioning, Navigation and Communication*. IEEE, March 2007. doi: 10.1109/wpnc.2007.353618. 1

[15] Enrique Garcia, Pablo Poudereux, Alvaro Hernandez, Jesus Urena, and David Gualda. A robust uwb indoor positioning system for highly complex environments. In *2015 IEEE International Conference on Industrial Technology (ICIT)*. IEEE, March 2015. doi: 10.1109/icit.2015.7125601.

[16] Janis Tiemann, Florian Schweikowski, and Christian Wietfeld. Design of an uwb indoor-positioning system for uav navigation in gnss-denied environments. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, October 2015. doi: 10.1109/ipin.2015.7346960. 1

[17] Matteo Ridolfi, Stef Vandermeeren, Jense Defraye, Heidi Steendam, Joeri Gerlo, Dirk De Clercq, Jeroen Hoebeke, and Eli De Poorter. Experimental evaluation of UWB indoor positioning for sport postures. *Sensors*, 18(2):168, jan 2018. doi: 10.3390/s18010168. 1, 76

[18] Zhang Sheng, Wen-De Zhong, Du Pengfei, Chen Chen, and Dehao Wu. Pdoa based indoor visible light positioning system without local oscillators in receiver. In *2017 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*. IEEE, July 2017. doi: 10.1109/cleopr.2017.8119047. 2

[19] Sheng Zhang, Wen-De Zhong, Pengfei Du, and Chen Chen. Experimental demonstration of indoor sub-decimeter accuracy vlp system using differential pdoa. *IEEE Photonics Technology Letters*, 30(19):1703–1706, October 2018. ISSN 1941-0174. doi: 10.1109/lpt.2018.2866402.

[20] Sheng Zhang, Pengfei Du, Chen Chen, Wen-De Zhong, and Arokiaswami Alphones. Robust 3d indoor vlp system based on ann using hybrid rss/pdoa. *IEEE Access*, 7:47769–47780, 2019. ISSN 2169-3536. doi: 10.1109/access.2019.2909761. 16, 22

[21] Helin Yang, Chen Chen, Wen-De Zhong, Sheng Zhang, and Pengfei Du. An integrated indoor visible light communication and positioning system based on fbmc-scm. In *2017 IEEE Photonics Conference (IPC)*. IEEE, October 2017. doi: 10.1109/ipcon.2017.8116035. 2

[22] Lin Bai, Yang Yang, Mingzhe Chen, Chunyan Feng, Caili Guo, Walid Saad, and Shuguang Cui. Computer vision-based localization with visible light communications. *IEEE Transactions on Wireless Communications*, 21(3):2051–2065, mar 2022. doi: 10.1109/twc.2021.3109146. 2, 19, 23, 25, 26, 30, 31, 32, 38, 46, 49

[23] Joon-Woo Lee, Sung-Jin Kim, and Sang-Kook Han. 3d visible light indoor positioning by bokeh based optical intensity measurement in smartphone camera. *IEEE Access*, 7:91399–91406, 2019. ISSN 2169-3536. doi: 10.1109/access.2019.2927356. 15

[24] Wansheng Pan, Yinan Hou, and Shilin Xiao. Visible light indoor positioning based on camera with specular reflection cancellation. In *2017 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*. IEEE, July 2017. doi: 10.1109/cleopr.2017.8118636.

[25] Ran Zhang, Wen-De Zhong, and Qian Kemao. A singular value decomposition-based positioning algorithm for indoor visible light positioning system. In *2017 Conference on Lasers and Electro-Optics Pacific Rim (CLEO-PR)*. IEEE, July 2017. doi: 10.1109/cleopr.2017.8118948. 2, 18

[26] Harald Haas. High-speed wireless networking using visible light. *SPIE Newsroom*, April 2013. ISSN 1818-2259. doi: 10.1117/2.1201304.004773. 2

[27] Wenhua Shao, Haiyong Luo, Fang Zhao, Yan Ma, Zhongliang Zhao, and Antonino Crivello. Indoor positioning based on fingerprint-image and deep learning. *IEEE Access*, 6:74699–74712, 2018. doi: 10.1109/access.2018.2884193. 2, 25, 55, 105

[28] Stéphane Beauregard and Harald Haas. Pedestrian dead reckoning: A basis for personal positioning. *Proceedings of the 3rd Workshop on Positioning, Navigation and Communication (WPNC'06)*, 01 2006. 2

[29] Xu Feng, Khuong An Nguyen, and Zhiyuan Luo. A survey of deep learning approaches for WiFi-based indoor positioning. *Journal of Information and Telecommunication*, 6(2):163–216, sep 2021. doi: 10.1080/24751839.2021.1975425. 2, 23, 76, 93

[30] Ran Zhang, Wen-De Zhong, Qian Kemao, and Sheng Zhang. A single LED positioning system based on circle projection. *IEEE Photonics Journal*, 9(4):1–9, aug 2017. doi: 10.1109/jphot.2017.2722474. 3, 19, 23, 25, 26, 31, 38, 44

[31] Yuan Zhuang, Luchi Hua, Longning Qi, Jun Yang, Pan Cao, Yue Cao, Yongpeng Wu, John Thompson, and Harald Haas. A survey of positioning systems using visible led lights. *IEEE Communications Surveys and Tutorials*, 20(3):1963–1988, 2018. ISSN 2373-745X. doi: 10.1109/comst.2018.2806558. 3, 14, 16, 23

[32] Fu-Kwun Wang and Yi-Chen Lu. Useful lifetime analysis for high-power white leds. *Microelectronics Reliability*, 54(6–7):1307–1315, June 2014. ISSN 0026-2714. doi: 10.1016/j.microrel.2014.02.029. 4

[33] Musa Furkan Keskin, Ahmet Dundar Sezer, and Sinan Gezici. Localization via visible light systems. *Proceedings of the IEEE*, 106(6):1063–1088, June 2018. ISSN 1558-2256. doi: 10.1109/jproc.2018.2823500. 14

[34] Weizhi Zhang, M. I. Sakib Chowdhury, and Mohsen Kavehrad. Asynchronous indoor positioning system based on visible light communications. *Optical Engineering*, 53(4):045105, April 2014. ISSN 0091-3286. doi: 10.1117/1.oe.53.4.045105. 15

[35] S.-H. Yang, E.-M. Jeong, D.-R. Kim, H.-S. Kim, Y.-H. Son, and S.-K. Han. Indoor three-dimensional location estimation based on led visible light communication. *Electronics Letters*, 49(1):54–56, January 2013. ISSN 1350-911X. doi: 10.1049/el.2012.3167. 15

[36] Hyun-Seung Kim, Deok-Rae Kim, Se-Hoon Yang, Yong-Hwan Son, and Sang-Kook Han. An indoor visible light communication positioning system using a rf carrier allocation technique. *Journal of Lightwave Technology*, 31(1):134–144, January 2013. ISSN 1558-2213. doi: 10.1109/jlt.2012.2225826. 15

[37] Thomas Q. Wang, Y. Ahmet Sekercioglu, Adrian Neild, and Jean Armstrong. Position accuracy of time-of-arrival based ranging using visible light with application in indoor localization systems. *Journal of Lightwave Technology*, 31(20):3302–3308, October 2013. ISSN 1558-2213. doi: 10.1109/jlt.2013.2281592. 15

[38] Soo-Yong Jung, Swook Hann, and Chang-Soo Park. Tdoa-based optical wireless indoor localization using led ceiling lamps. *IEEE Transactions on Consumer Electronics*, 57(4):1592–1597, November 2011. ISSN 0098-3063. doi: 10.1109/tce.2011.6131130. 15

[39] U. Nadeem, N.U. Hassan, M.A. Pasha, and C. Yuen. Highly accurate 3d wireless indoor positioning system using white led lights. *Electronics Letters*, 50(11):828–830, May 2014. ISSN 1350-911X. doi: 10.1049/el.2014.0353. 15

[40] Gregary B. Prince and Thomas D.C. Little. A two phase hybrid rss/aoa algorithm for indoor device localization using visible light. In *2012 IEEE Global Communications Conference (GLOBECOM)*. IEEE, December 2012. doi: 10.1109/glocom.2012.6503631. 16

[41] Yusuf Said Eroglu, Ismail Guvenc, Nezih Pala, and Murat Yuksel. Aoa-based localization and tracking in multi-element vlc systems. In *2015 IEEE 16th Annual*

Wireless and Microwave Technology Conference (WAMICON). IEEE, April 2015. doi: 10.1109/wamicon.2015.7120424. 16

[42] Chinnapat Sertthin, Emiko Tsuji, Masao Nakagawa, Shigeru Kuwano, and Kazuji Watanabe. A switching estimated receiver position scheme for visible light based indoor positioning system. In *2009 4th International Symposium on Wireless Pervasive Computing*. IEEE, February 2009. doi: 10.1109/iswpc.2009.4800561. 16

[43] Muhammad Yasir, Siu-Wai Ho, and Badri N. Vellambi. Indoor positioning system using visible light and accelerometer. *Journal of Lightwave Technology*, 32(19): 3306–3316, October 2014. ISSN 1558-2213. doi: 10.1109/jlt.2014.2344772. 16

[44] Georg Kail, Patrick Maechler, Nicholas Preyss, and Andreas Burg. Robust asynchronous indoor localization using led lighting. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, May 2014. doi: 10.1109/icassp.2014.6853922. 16

[45] Zhijie Luo, WeiNan Zhang, and GuoFu Zhou. Improved spring model-based collaborative indoor visible light positioning. *Optical Review*, 23(3):479–486, March 2016. ISSN 1349-9432. doi: 10.1007/s10043-016-0204-z. 16

[46] Huy Q. Tran and Cheolkeun Ha. Reducing the burden of data collection in a fingerprinting-based vlp system using a hybrid of improved co-training semi-supervised regression and adaptive boosting algorithms. *Optics Communications*, 488:126857, June 2021. ISSN 0030-4018. doi: 10.1016/j.optcom.2021.126857. 16

[47] Ran Zhang, Wen-De Zhong, Dehao Wu, and Kemao Qian. A novel sensor fusion based indoor visible light positioning system. In *2016 IEEE Globecom Workshops (GC Wkshps)*. IEEE, December 2016. doi: 10.1109/glocomw.2016.7848823. 18, 23

[48] Ran Zhang, Wen-De Zhong, Ke-mao Qian, and De-hao Wu. Image sensor based visible light positioning system with improved positioning algorithm. *IEEE Access*, pages 1–1, 2017. ISSN 2169-3536. doi: 10.1109/access.2017.2693299.

[49] Jiaojiao Xu, Chen Gong, and Zhengyuan Xu. Indoor visible light positioning with centimeter accuracy based on a commercial smartphone camera. In *2018 IEEE Globecom Workshops (GC Wkshps)*. IEEE, December 2018. doi: 10.1109/glocomw.2018.8644462. 18

[50] Yoli Shavit and Ron Ferens. Introduction to camera pose estimation with deep learning, 2019. 20

[51] Alex Kendall, Matthew Grimes, and Roberto Cipolla. Posenet: A convolutional network for real-time 6-dof camera relocalization. In *2015 IEEE International Conference on Computer Vision (ICCV)*. IEEE, December 2015. doi: 10.1109/iccv.2015.336. 20, 22

[52] Y. Bengio, A. Courville, and P. Vincent. Representation learning: A review and new perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, August 2013. ISSN 2160-9292. doi: 10.1109/tpami.2013.50. 20

[53] Jeff Donahue, Yangqing Jia, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, and Trevor Darrell. Decaf: A deep convolutional activation feature for generic visual recognition, 2013.

[54] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. Cnn features off-the-shelf: an astounding baseline for recognition, 2014.

[55] Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. Learning and transferring mid-level image representations using convolutional neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2014. doi: 10.1109/cvpr.2014.222. 20

[56] Alexander Toshev and Christian Szegedy. Deeppose: Human pose estimation via deep neural networks. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2014. doi: 10.1109/cvpr.2014.214. 20

[57] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions, 2014. 20, 21

[58] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. Efficient and effective prioritized matching for large-scale image-based localization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(9):1744–1756, September 2017. ISSN 2160-9292. doi: 10.1109/tpami.2016.2611662. 21

[59] Torsten Sattler, Bastian Leibe, and Leif Kobbelt. *Improving Image-Based Localization by Active Correspondence Search*, pages 752–765. Springer Berlin Heidelberg, 2012. ISBN 9783642337185. doi: 10.1007/978-3-642-33718-5_54. 21

[60] F. Walch, C. Hazirbas, L. Leal-Taixe, T. Sattler, S. Hilsenbeck, and D. Cremers. Image-based localization using lstms for structured feature correlation. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, October 2017. doi: 10.1109/iccv.2017.75. 21

[61] Abhinav Valada, Noha Radwan, and Wolfram Burgard. Deep auxiliary learning for visual localization and odometry. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, May 2018. doi: 10.1109/icra.2018.8462979. 21

[62] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. 21

[63] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, June 2009. doi: 10.1109/cvpr.2009.5206848. 21

[64] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, and Antonio Torralba. Places: A 10 million image database for scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(6):1452–1464, June 2018. ISSN 1939-3539. doi: 10.1109/tpami.2017.2723009. 21

[65] Xiangyu Yue, Bichen Wu, Sanjit A. Seshia, Kurt Keutzer, and Alberto L. Sangiovanni-Vincentelli. A lidar point cloud generator: from a virtual world to autonomous driving. In *Proceedings of the 2018 ACM on International Conference on Multimedia Retrieval*, ICMR '18. ACM, June 2018. doi: 10.1145/3206025.3206080. 22

[66] A Geiger, P Lenz, C Stiller, and R Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, August 2013. ISSN 1741-3176. doi: 10.1177/0278364913491297. 22

[67] Angel X. Chang, Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese, Manolis Savva, Shuran Song, Hao Su, Jianxiong Xiao, Li Yi, and Fisher Yu. Shapenet: An information-rich 3d model repository, 2015. 22

[68] Xudong Song, Xiaochen Fan, Chaocan Xiang, Qianwen Ye, Leyu Liu, Zumin Wang, Xiangjian He, Ning Yang, and Gengfa Fang. A novel convolutional neural network based indoor localization framework with WiFi fingerprinting. *IEEE Access*, 7:110698–110709, 2019. doi: 10.1109/access.2019.2933921. 23, 76, 91, 92, 93, 96, 101

[69] Letícia Fernandes, Sara Santos, Marília Barandas, Duarte Folgado, Ricardo Leonardo, Ricardo Santos, André Carreiro, and Hugo Gamboa. An infrastructure-free magnetic-based indoor positioning system with deep learning. *Sensors*, 20 (22):6664, November 2020. ISSN 1424-8220. doi: 10.3390/s20226664. 23

[70] Fabrizio De Vita and Dario Bruneo. A deep learning approach for indoor user localization in smart environments. In *2018 IEEE International Conference on Smart Computing (SMARTCOMP)*. IEEE, June 2018. doi: 10.1109/smartcomp.2018.00078.

[71] Fahad Al-homayani and Mohammad Mahoor. Improved indoor geomagnetic field fingerprinting for smartwatch localization using deep learning. In *2018 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, September 2018. doi: 10.1109/ipin.2018.8626558. 23, 24

[72] Leonid Antsfeld, Boris Chidlovskii, and Emilio Sansano-Sansano. Deep smartphone sensors-wifi fusion for indoor positioning and tracking, 2020. 23

[73] Wenxu Wang, Damián Marelli, and Minyue Fu. Fingerprinting-based indoor localization using interpolated preprocessed csi phases and bayesian tracking. *Sensors*, 20(10):2854, May 2020. ISSN 1424-8220. doi: 10.3390/s20102854. 23

[74] Joaquin Torres-Sospedra, Raul Montoliu, Adolfo Martinez-Uso, Joan P. Avariento, Tomas J. Arnau, Mauri Benedito-Bordonau, and Joaquin Huerta. UJIIndoorLoc: A new multi-building and multi-floor database for WLAN fingerprint-based indoor localization problems. In *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, oct 2014. doi: 10.1109/ipin.2014.7275492. 24, 76, 78, 84, 101

[75] Elena Lohan, Joaquín Torres-Sospedra, Helena Leppäkoski, Philipp Richter, Zhe Peng, and Joaquín Huerta. Wi-fi crowdsourced fingerprinting dataset for indoor positioning. *Data*, 2(4):32, oct 2017. doi: 10.3390/data2040032. 24, 101

[76] M. Stella, M. Russo, and D. Begusic. Location determination in indoor environment based on rss fingerprinting and artificial neural network. In *2007 9th International Conference on Telecommunications*. IEEE, June 2007. doi: 10.1109/contel.2007.381886. 24

[77] Huan Dai, Wen-hao Ying, and Jiang Xu. Multi-layer neural network for received signal strength-based indoor localisation. *IET Communications*, 10(6):717–723, April 2016. ISSN 1751-8636. doi: 10.1049/iet-com.2015.0469. 24

[78] Joaquin Torres-Sospedra, Darwin P. Quezada Gaibor, Jari Nurmi, Yevgeni Koucheryavy, Elena Simona Lohan, and Joaquin Huerta. Scalable and efficient clustering for fingerprint-based positioning. *IEEE Internet of Things Journal*, 10(4):3484–3499, feb 2023. doi: 10.1109/jiot.2022.3230913. 25

[79] Srivathsan Chakaravarthi Narasimman and Arokiaswami Alphones. Dumbloc: Dumb indoor localization framework using wifi fingerprinting. *IEEE Sensors Journal*, pages 1–1, 2024. ISSN 2379-9153. doi: 10.1109/jsen.2024.3374415. 25, 55

[80] Loizos Kanaris, Akis Kokkinis, Antonio Liotta, and Stavros Stavrou. Combining smart lighting and radio fingerprinting for improved indoor localization. In *2017 IEEE 14th International Conference on Networking, Sensing and Control (ICNSC)*. IEEE, may 2017. doi: 10.1109/icnsc.2017.8000134. 25, 37, 76, 95

[81] Zhiyu Zhu, Yang Yang, Mingzhe Chen, Caili Guo, and Yipeng Bai. Visible light positioning based on a single luminaire: A novel visual odometry assisted algorithm. In *ICC 2023 - IEEE International Conference on Communications*. IEEE, May 2023. doi: 10.1109/icc45041.2023.10279121. 25, 26, 31

[82] Han Cheng, Chunxian Xiao, Yongqing Ji, Jianmin Ni, and Tongyao Wang. A single led visible light positioning system based on geometric features and cmos camera. *IEEE Photonics Technology Letters*, 32(17):1097–1100, September 2020. ISSN 1941-0174. doi: 10.1109/lpt.2020.3012476. 26, 31

[83] Jie Hao, Jing Chen, and Ran Wang. Visible light positioning using a single led luminaire. *IEEE Photonics Journal*, 11(5):1–13, October 2019. ISSN 1943-0647. doi: 10.1109/jphot.2019.2930209. 26, 31

[84] Babar Hussain, Yiru Wang, Runzhou Chen, and C. Patrick Yue. Camera pose estimation using a vlc-modulated single rectangular led for indoor positioning.

IEEE Transactions on Instrumentation and Measurement, 71:1–11, 2022. ISSN 1557-9662. doi: 10.1109/tim.2022.3212980. 26, 31

[85] Srivathsan Chakaravarthi Narasimman and Arokiaswami Alphones. Tree-based single led indoor visible light positioning technique. In *TENCON 2023 - 2023 IEEE Region 10 Conference (TENCON)*. IEEE, October 2023. doi: 10.1109/tencon58879.2023.10322350. 26, 30, 31, 55, 76

[86] Luchi Hua, Yuan Zhuang, and Jun Yang. Deep learning-based fusion of visible light positioning and IMU sensors. In *2021 20th International Conference on Ubiquitous Computing and Communications (IUCC/CIT/DSCI/SmartCNS)*. IEEE, dec 2021. doi: 10.1109/iucc-cit-dsci-smartcns55181.2021.00093. 26

[87] Chao Qin and Xingqun Zhan. VLIP: Tightly coupled visible-light/inertial positioning system to cope with intermittent outage. *IEEE Photonics Technology Letters*, 31(2):129–132, jan 2019. doi: 10.1109/lpt.2018.2883345. 26

[88] Jianbo Shi and Tomasi. Good features to track. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition CVPR-94*. IEEE Comput. Soc. Press, 1994. doi: 10.1109/cvpr.1994.323794. 29, 41

[89] Ye-Sheng Kuo, Pat Pannuto, Ko-Jen Hsiao, and Prabal Dutta. Luxapose. In *Proceedings of the 20th annual international conference on Mobile computing and networking*. ACM, sep 2014. doi: 10.1145/2639108.2639109. 29, 56

[90] Weipeng Guan, Yuxiang Wu, Canyu Xie, Liangtao Fang, Xiaowei Liu, and Yingcong Chen. Performance analysis and enhancement for visible light communication using CMOS sensors. *Optics Communications*, 410:531–551, mar 2018. doi: 10.1016/j.optcom.2017.10.038. 37, 39

[91] Tao Yuan, Yiqin Xu, Yong Wang, Peng Han, and Junfang Chen. A tilt receiver correction method for visible light positioning using machine learning method. *IEEE Photonics Journal*, 10(6):1–12, dec 2018. doi: 10.1109/jphot.2018.2880872. 38

[92] Peixi Liu, Tianqi Mao, Ke Ma, Jiaxuan Chen, and Zhaocheng Wang. Three-dimensional visible light positioning using regression neural network. In *2019 15th International Wireless Communications &Mobile Computing Conference (IWCMC)*. IEEE, jun 2019. doi: 10.1109/iwcmc.2019.8766658. 38

[93] Juan D. Gutierrez, Teodoro Aguilera, Fernando J. Alvarez, Jorge Morera, and Fernando J. Aranda. A blender-based simulation tool for visible light positioning with portable devices. In *2022 IEEE International Instrumentation and Measurement Technology Conference (I2MTC)*. IEEE, may 2022. doi: 10.1109/i2mtc48687.2022.9806547. 38

[94] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. URL `@Manual{Community2018, title={Blender-a3Dmodellingandrenderingpackage}, address={StichtingBlenderFoundation,Amsterdam},author={BlenderOnlineCommunity},organization={BlenderFoundation}, year={2018},url={http://www.blender.org},}`. 38

[95] Chang Bek Mei and lee Cliff. *BCA-circular for BIM e-submission requirements for as-built plan submissions to BCA*. Building and Construction Authority, 2020. URL `https://www1.bca.gov.sg/docs/default-source/docs-corp-news-and-publications/circulars/bim-circular-for-private-project_dec2020.pdf`. 44

[96] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/a:1010933404324. 44, 97

[97] Tianqi Chen and Carlos Guestrin. XGBoost. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, aug 2016. doi: 10.1145/2939672.2939785. 44, 92

[98] Simon Haykin. *Neural networks: a comprehensive foundation*. Prentice Hall PTR, 1994. 44

[99] Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Andreas Müller, Joel Nothman, Gilles Louppe, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. 2012. doi: 10.48550/ARXIV.1201.0490. 45, 87, 92, 99

[100] Léo Grinsztajn, Edouard Oyallon, and Gaël Varoquaux. Why do tree-based models still outperform deep learning on tabular data?, 2022. 45, 92, 93

[101] Chi-Wai Chow, Chung-Yen Chen, and Shih-Hao Chen. Enhancement of signal performance in led visible light communications using mobile phone camera. *IEEE Photonics Journal*, 7(5):1–7, October 2015. ISSN 1943-0655. doi: 10.1109/jphot.2015.2476757. 55

[102] Srivathsan Chakaravarthi Narasimman and Arokiaswami Alphones. Indoor visible light positioning for a single partially visible led. *IEEE Sensors Letters*, 8(5):1–4, May 2024. ISSN 2475-1472. doi: 10.1109/lsens.2024.3385543. 55

[103] Yuki Goto, Isamu Takai, Takaya Yamazato, Hiraku Okada, Toshiaki Fujii, Shoji Kawahito, Shintaro Arai, Tomohiro Yendo, and Koji Kamakura. A new automotive vlc system using optical communication image sensor. *IEEE Photonics Journal*, 8(3):1–17, June 2016. ISSN 1943-0655. doi: 10.1109/jphot.2016.2555582. 55

[104] Pengfei Luo, Min Zhang, Zabih Ghassemlooy, Hoa Le Minh, Hsin-Mu Tsai, Xuan Tang, Lih Chieh Png, and Dahai Han. Experimental demonstration of rgb led-based optical camera communications. *IEEE Photonics Journal*, 7(5):1–12, October 2015. ISSN 1943-0655. doi: 10.1109/jphot.2015.2486680. 55

[105] Joyce E. Farrell, Peter B. Catrysse, and Brian A. Wandell. Digital camera simulation. *Applied Optics*, 51(4):A80, February 2012. ISSN 2155-3165. doi: 10.1364/ao.51.000a80. 55

[106] Anqi Liu, Wenxiao Shi, Wei Liu, and Zhuo Wang. A simplified system model for optical camera communication. In *2021 IEEE/CIC International Conference on Communications in China (ICCC)*. IEEE, July 2021. doi: 10.1109/iccc52777.2021.9580326. 55

[107] Alexis Duque, Razvan Stanica, Herve Rivano, and Adrien Desportes. Analytical and simulation tools for optical camera communications. *Computer Communications*, 160:52–62, July 2020. ISSN 0140-3664. doi: 10.1016/j.comcom.2020.05.036. 56

[108] Moh. Khalid Hasan, Mostafa Zaman Chowdhury, Md. Shahjalal, Van Thang Nguyen, and Yeong Min Jang. Performance analysis and improvement of optical camera communication. *Applied Sciences*, 8(12):2527, December 2018. ISSN 2076-3417. doi: 10.3390/app8122527. 56, 57

[109] Trong-Hop Do and Myungsik Yoo. Performance analysis of visible light communication using cmos sensors. *Sensors*, 16(3):309, February 2016. ISSN 1424-8220. doi: 10.3390/s16030309. 56, 59, 62

[110] Yang Liu, Kevin Liang, Hung-Yu Chen, Liang-Yu Wei, Chin-Wei Hsu, Chi-Wai Chow, and Chien-Hung Yeh. Light encryption scheme using light-emitting diode and camera image sensor. *IEEE Photonics Journal*, 8(1):1–7, February 2016. ISSN 1943-0655. doi: 10.1109/jphot.2016.2519287. 56, 60, 66

[111] *Handbook of Photoelectric Sensing*. Banner Engineering Corp., Banner Engineering: Plymouth, MN, USA, second edition, 1993. 59

[112] Jin Shi, Jing He, and Xinda Yan. Sub-column pixel neural network scheme for modulation format shifting based optical camera communications. *Optics Letters*, 48(1):85, December 2022. ISSN 1539-4794. doi: 10.1364/ol.479716. 61

[113] Jing He, Yiting Yang, and Jing He. Artificial neural network-based scheme for 4-pwm occ system. *IEEE Photonics Technology Letters*, 34(6):333–336, March 2022. ISSN 1941-0174. doi: 10.1109/lpt.2022.3153692. 61

[114] Kevin Buchin, Maike Buchin, Wouter Meulemans, and Bettina Speckmann. Locally correct fréchet matchings. *Computational Geometry*, 76:1–18, January 2019. ISSN 0925-7721. doi: 10.1016/j.comgeo.2018.09.002. 63

[115] Sebastian Sadowski and Petros Spachos. RSSI-based indoor localization with the internet of things. *IEEE Access*, 6:30149–30161, 2018. doi: 10.1109/access.2018.2843325. 76

[116] Qianwen Ye, Xiaochen Fan, Gengfa Fang, Hongxia Bie, Xudong Song, and Rajan Shankaran. CapsLoc: A robust indoor localization system with WiFi fingerprinting using capsule networks. In *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*. IEEE, jun 2020. doi: 10.1109/icc40277.2020.9148933. 76, 92

[117] Roman Klus, Lucie Klus, Jukka Talvitie, Jaakko Pihlajasalo, Joaquin Torres-Sospedra, and Mikko Valkama. Transfer learning for convolutional indoor positioning systems. In *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, nov 2021. doi: 10.1109/ipin51156.2021.9662544. 76

[118] Y. Oh, H.-M. Noh, and W. Shin. C-CNNLoc: Constrained CNN for robust indoor localization with building boundary. *Electronics Letters*, 57(10):422–425, mar 2021. doi: 10.1049/ell2.12142. 76

[119] Aleksandr Ometov, Viktoriia Shubina, Lucie Klus, Justyna Skibińska, Salwa Saafi, Pavel Pascacio, Laura Flueratoru, Darwin Quezada Gaibor, Nadezhda Chukhno, Olga Chukhno, Asad Ali, Asma Channa, Ekaterina Svertoka, Waleed Bin Qaim, Raúl Casanova-Marqués, Sylvia Holcer, Joaquín Torres-Sospedra, Sven Casteleyn, Giuseppe Ruggeri, Giuseppe Araniti, Radim Burget, Jiri Hosek, and Elena Simona Lohan. A survey on wearable technology: History, state-of-the-art and current challenges. *Computer Networks*, 193:108074, jul 2021. doi: 10.1016/j.comnet.2021.108074. 76

[120] Alwin Poulose and Dong Seog Han. Indoor localization using PDR with wi-fi weighted path loss algorithm. In *2019 International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, oct 2019. doi: 10.1109/ictc46691.2019.8939753. 76, 95

[121] Rizanne Elbakly and Moustafa Youssef. The StoryTeller. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 4(1):1–20, mar 2020. doi: 10.1145/3380979. 76

[122] Marius Laska and Jorg Blankenbach. Topology preserving input image for convolutional neural network based indoor localization. In *2021 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, nov 2021. doi: 10.1109/ipin51156.2021.9662471. 76

[123] Jun Gao, Ceyao Zhang, Qinglei Kong, Feng Yin, Lexi Xu, and Kai Niu. MetaLoc: Learning to learn indoor RSS fingerprinting localization over multiple scenarios. In *ICC 2022 - IEEE International Conference on Communications*. IEEE, may 2022. doi: 10.1109/icc45855.2022.9838587. 76, 92

[124] Germán Mendoza-Silva, Philipp Richter, Joaquín Torres-Sospedra, Elena Lohan, and Joaquín Huerta. Long-term WiFi fingerprinting dataset for research on robust indoor positioning. *Data*, 3(1):3, jan 2018. doi: 10.3390/data3010003. 78, 101

[125] N. Bulusu, J. Heidemann, and D. Estrin. GPS-less low-cost outdoor localization for very small devices. *IEEE Personal Communications*, 7(5):28–34, 2000. doi: 10.1109/98.878533. 79

[126] Christine Laurendeau and Michel Barbeau. Centroid localization of uncooperative nodes in wireless networks using a relative span weighting method. *EURASIP Journal on Wireless Communications and Networking*, 2010(1), nov 2009. doi: 10.1155/2010/567040. 80, 87

[127] Yu-Chung Cheng, Yatin Chawathe, Anthony LaMarca, and John Krumm. Accuracy characterization for metropolitan-scale wi-fi localization. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services*. ACM, jun 2005. doi: 10.1145/1067170.1067195. 80

[128] Jan Blumenthal, Ralf Grossmann, Frank Golatowski, and Dirk Timmermann. Weighted centroid localization in zigbee-based sensor networks. In *2007 IEEE International Symposium on Intelligent Signal Processing*. IEEE, 2007. doi: 10.1109/wisp.2007.4447528. 80, 87

[129] Henri Nurminen, Marzieh Dashti, and Robert Piché. A survey on wireless transmitter localization using signal strength measurements. *Wireless Communications and Mobile Computing*, 2017:1–12, 2017. doi: 10.1155/2017/2569645. 80, 87

[130] Theodore S. Rappaport. *Wireless Communications: Principles and Practice*. IEEE. ISBN 9780780311671. 80

[131] Ananth Ranganathan. The levenberg-marquardt algorithm. *Tutoral on LM algorithm*, 11(1):101–110, 2004. 81

[132] K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2):182–197, apr 2002. doi: 10.1109/4235.996017. 81

[133] Mohd Amiruddin Abd Rahman, Marzieh Dashti, and Jie Zhang. Localization of unknown indoor wireless transmitter. In *2013 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, jun 2013. doi: 10.1109/icl-gnss.2013.6577270. 81

[134] Henri Nurminen, Jukka Talvitie, Simo Ali-Loytty, Philipp Muller, Elena-Simona Lohan, Robert Piche, and Markku Renfors. Statistical path loss parameter estimation and positioning using RSS measurements in indoor wireless networks. In *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, nov 2012. doi: 10.1109/ipin.2012.6418856. 81

[135] Rizanne Elbakly, Heba Aly, and Moustafa Youssef. TrueStory: Accurate and robust RF-based floor estimation for challenging indoor environments. *IEEE Sensors Journal*, 18(24):10115–10124, dec 2018. doi: 10.1109/jsen.2018.2872827. 86

[136] Aina Nadhirah Nor Hisham, Yin Hoe Ng, Chee Keong Tan, and David Chieng. Hybrid wi-fi and BLE fingerprinting dataset for multi-floor indoor environments with different layouts. *Data*, 7(11):156, nov 2022. doi: 10.3390/data7110156. 87

[137] Joaquín Torres-Sospedra, Raúl Montoliu, Sergio Trilles, Óscar Belmonte, and Joaquín Huerta. Comprehensive analysis of distance and similarity measures for wi-fi fingerprinting indoor positioning systems. *Expert Systems with Applications*, 42(23):9263–9278, dec 2015. doi: 10.1016/j.eswa.2015.08.013. 91

[138] Adriano Moreira, Maria Joao Nicolau, Filipe Meneses, and Antonio Costa. Wi-fi fingerprinting in the real world - RTLS@UM at the EvAAL competition. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*. IEEE, oct 2015. doi: 10.1109/ipin.2015.7346967. 96

[139] Marius Laska and Jörg Blankenbach. DeepLocBox: Reliable fingerprinting-based indoor area localization. *Sensors*, 21(6):2000, mar 2021. doi: 10.3390/s21062000. 96

[140] Erwan Scornet. Trees, forests, and impurity-based variable importance. January 2020. doi: 10.48550/ARXIV.2001.04295. 97

[141] Wei-Yin Loh. Classification and regression trees. *WIREs Data Mining and Knowledge Discovery*, 1(1):14–23, jan 2011. doi: 10.1002/widm.8. 97

[142] L. Breiman. Manual on setting up, using, and understanding random forests v3. 1. Technical Report 1:58, Statistics Department University of California Berkeley, CA, USA, 2002. 99

[143] Elena Simona Lohan, Joaquín Torres-Sospedra, and Alejandro Gonzalez. Wifi rss measurements in tampere university multi-building campus, 2017, 2021. URL `https://zenodo.org/record/5174851`. 101

[144] Shweta Shrestha, Jukka Talvitie, and Elena Simona Lohan. Deconvolution-based indoor localization with WLAN signals and unknown access point locations. In

2013 International Conference on Localization and GNSS (ICL-GNSS). IEEE, jun 2013. doi: 10.1109/icl-gnss.2013.6577256. 101

[145] Philipp Richter, Elena Simona Lohan, and Jukka Talvitie. Wlan (wifi) rss database for fingerprinting positioning, 2018. URL `https://zenodo.org/record/1161525`. 101

[146] Lucie Klus, Darwin Quezada-Gaibor, Joaquin Torres-Sospedra, Elena Simona Lohan, Carlos Granell, and Jari Nurmi. Towards accelerated localization performance across indoor positioning datasets. In *2022 International Conference on Localization and GNSS (ICL-GNSS)*. IEEE, jun 2022. doi: 10.1109/icl-gnss54081.2022.9797035. 103, 106

[147] Huy Nguyen, Ida Bagus Krishna Yoga Utama, and Yeong Min Jang. Enabling technologies and new challenges in ieee 802.15.7 optical camera communications standard. *IEEE Communications Magazine*, 62(3):90–95, March 2024. ISSN 1558-1896. doi: 10.1109/mcom.002.2300289. 116

[148] Bekir Sait Ciftler, Abdullatif Albaseer, Noureddine Lasla, and Mohamed Abdallah. Federated learning for localization: A privacy-preserving crowdsourcing method, 2020. 117

[149] Zheshun Wu, Xiaoping Wu, and Yunliang Long. Prediction based semi-supervised online personalized federated learning for indoor localization. *IEEE Sensors Journal*, 22(11):10640–10654, June 2022. ISSN 2379-9153. doi: 10.1109/jsen.2022.3165042. 117