Spring 5-2-2022

Improved Visible Light Communication Receiver Performance by Leveraging the Spatial Dimension

MD Rashed Rahman

doi: https://doi.org/10.57709/28902536

IMPROVED VISIBLE LIGHT COMMUNICATION RECEIVER PERFORMANCE BY LEVERAGING THE SPATIAL DIMENSION

by

MD RASHED RAHMAN

Under the Direction of Ashwin Ashok, PhD

ABSTRACT

In wireless communications systems, signals can be transmitted as time (temporal) or spatial variants across 3D space, and in both ways. However, using temporal variant communication channels in high-speed data transmission introduces inter-symbol interference (ISI) which makes the systems unreliable. On the other hand, spatial diversity in signal processing reduces the ISI and improves the system throughput or performance by allowing more signals from different spatial locations at the same time. Therefore, the spatial features or properties of visible light signals can be very useful in designing a reliable visible light communication (VLC) system with higher system throughput and making it more robust against ambient noise and interference. By allowing only the signals of interest, spatial separability in VLC

can minimize the noise to a greater extent to improve signal-to-noise ratio (SNR) which can ensure higher data rates (in the order of Gbps-Tbps) in VLC. So, designing a VLC system with spatial diversity is an exciting area to explore and might set the foundation for future VLC system architectures and enable different VLC based applications such as vehicular VLC, multi-VLC, localization, and detection using VLC, etc. This thesis work is motivated by the fundamental challenges in reusing spatial information in VLC systems to increase the system throughput or gain through novel system designing and their prototype implementations.

INDEX WORDS: Visible Light Communication, Optical Wireless Sensing and Localization, LED-Camera Communication, Spatial Signal Processing, High Speed VLC, Multiple Access, Vehicular VLC.

IMPROVED VISIBLE LIGHT COMMUNICATION RECEIVER PERFORMANCE BY LEVERAGING THE SPATIAL DIMENSION

by

MD RASHED RAHMAN

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2022

IMPROVED VISIBLE LIGHT COMMUNICATION RECEIVER PERFORMANCE BY LEVERAGING THE SPATIAL DIMENSION

by

MD RASHED RAHMAN

Committee Chair:	Ashwin Ashok
Committee:	Rajshekhar Sunderraman
	Xiaojun Cao
	Murat Yuksel

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2022

DEDICATION

I want to dedicate my PhD dissertation to my parents (Amma and Abba), my sister (Apa), and my wife, Era, who never doubted my capabilities and objectives, and continuously supported me in every possible way.

ACKNOWLEDGMENTS

First of all, I would like to express my sincere gratitude towards my parents who sacrificed a lot to shape my future. Whatever success I have achieved or will be achieving in my life, their hard work and sacrifice would be the foundation of those successes. Both of my parents don't have formal education, but that didn't stop them from encouraging me to pursue the highest level of formal education. I would also like to thank my sister, who always supported me in whatever decisions I have taken in my life, just like my parents. I would be the first person in my family to have a PhD degree, which will surely make my family proud of me. I also like to express my heartfelt gratitude towards my beloved wife, Era, who is pursuing the same dream of being a doctorate while enduring all of the struggles of life by staying beside me. I cannot thank her enough for her unconditional support and encouragement throughout this long and challenging journey.

I am incredibly grateful to my supervisor, Dr. Ashwin Ashok, for allowing me to join his group and guiding me throughout my doctoral research journey with patience. It's my pleasure to meet Dr. Ashok, and the learning I got from him significantly contributed to my professional and personal growth. I also want to thank my other committee members, Dr. Raj Sunderraman, Dr. Xiaojun Cao, and Dr. Murat Yuksel, for their continuous support and guidance throughout my PhD journey. Finally, I appreciate all of my friends, fellow PhD students, faculties, and staff for their help and inspiration during my tenure at the GSU CS department for the last five years.

FUNDING ACKNOWLEDGMENT

The research projects for this dissertation have been supported by the National Science Foundation (NSF) under the grant NSF CRII CNS-NeTS 1755925, REU Supplement grant CNS-192917, and CNS Core-1901133.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

VLC : Visible Light Communication

LED : Light Emitting Diode

PD : Photodiode

ISI: Inter-Symbol Interference

V2V : Vehicle to Vehicle

Gbps: Giga Bits Per Second

Tbps: Tera Bits Per Second

LOS : Line-of-Sight

NLOS : Non-Line-of-Sight

SNR : Signal to Noise Ratio

SINR: Signal to Interference and Noise Ratio

BER : Bit Error Rate

PER : Packet Error Rate

FOV : Field-of-View

OFOV : Out of Field-of-View

LIDAR: Light Detection and Ranging

CHAPTER 1
INTRODUCTION

1.1 Visible Light Communication (VLC)

The significant growth in the wireless data traffic has initiated the need for expanding the range of frequencies used for wireless data communication. This has opened up new opportunities for utilizing the unused bands of the electromagnetic spectrum such as optical frequencies for wireless data communication through the Visible Light Communication (VLC) technology [2; 3]. VLC is a wireless communication technology that operates unregulated in the visible–light band (400–800 THz frequencies or 380–780 nm wavelengths) of the electromagnetic spectrum, and is enabled by light emitting elements such as light emitting diodes (LED) and light receiving elements such as photodiodes (PD).

The semiconductor properties of LEDs and PDs enable them to be switched at extremely high rates thus allowing transmission/reception of light beams at extremely high frequencies. VLC is a line-of-sight (LOS) technology, which means it requires the light transmitter and receiver to be within the distance and angular range (field–of–view (FOV)) of one another. The LOS requirement enables efficient space reuse allowing spatial–multiplexing of VLC links between multiple transmitters and receivers. The availability of a huge unrestricted visible-light spectrum and the spatio-temporal qualities makes VLC a strong proponent for high–speed wireless communication.

Over the past few years, VLC technology has garnered significant interest in both academic and industrial fronts. Research and development in VLC has exemplified VLC applica-

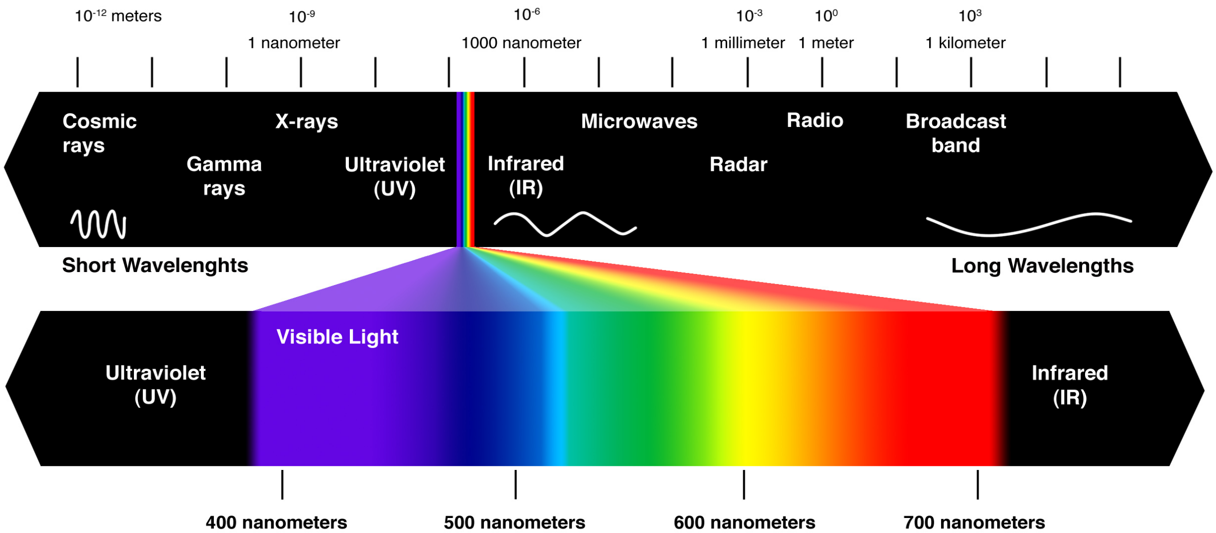


Figure 1.1: Visible light wavelength band in electromagnetic spectrum [1].

tions across diverse areas including smart sensing for human-computer interaction [4], precise indoor localization [5], inter-vehicular and vehicular to infrastructure communication [2] and underwater communication [6]. Operating over an unrestricted 400THz of bandwidth, VLC is capable of extremely high data rate communication, of the order of Gbps and beyond. VLC channel studies [7; 8; 9] estimate its data capacity to the order of Tbps. However, in practice, VLC systems are still operating in the range of Kbps–Mbps. So, there is a large gap to fill to reach ultra–high data rates or system throughput in VLC. Mobility is another huge challenge for VLC as optical links are highly directional and thus even the slightest movements of the transmitter and/or receiver can significantly degrade the link quality. The use of VLC as a next generation mobile wireless technology can be justified only if it can offer mobility in addition to high data communication speeds. Handling mobility can incur communication and processing overheads which can significantly degrade the VLC link data rate. Therefore, from an VLC architecture design standpoint, mobility and high data rate solutions have to be developed together. Also, there has long been interest in enabling visual interactions with phone cameras [10; 11; 12], including to assist local communication by obtaining a security token with the camera. Using VLC technology in sensing and leveraging light's directionality to improve localization accuracy is still one of the open research areas in the wireless sensing community.

Figure 1.2: Illustration of a typical indoor VLC (left) and outdoor vehicular VLC system (right).

By definition, **Spatial Signal Processing** offers the fundamental mathematical models and the spatial information of the signals including their physical properties, sources of the signals, and also, the geometric locations of each of the sources[13; 14; 15]. Utilizing such spatial information in VLC can ensure the following benefits:

- **Accessibility:** By measuring signals at different spatial locations and then allowing multiple signals into the receiver, the overall VLC system performance can be improved substantially.

- **Spatial Filtering:** To identify the incoming signals and blocking most of the noise outside the directions of interest will increase the SINR of the system and leads ultra high data rates in VLC.

- **Spatial Resolvability:** helps to distinguish between all the incoming signals from various sources into the VLC receiver and such capability can enable multiple access in VLC system even using a single receiver.

- **Spatial Locality:** Using geometric properties of a spatial domain, it is very much possible to locate the sources of all the incoming light signals and can be applicable in different VLC based object localization applications.

While it is clear that spatial signal processing can surely augment the overall system throughput or performance, therefore, in this thesis, we aim to explore new VLC systems and their implementations using the spatial dimensions of light signals.

1.2 Scopes of the thesis

In the beginning of this thesis, we developed a novel pixelated shutter based VLC receiver which can ensure higher signal reception by canceling noise and interference based on spatial selections of incoming signals. Our designed novel VLC receiver can select the exact area over which the transmitted signal is detected on the pixelated shutter array. Through this single photodiode (PD) based VLC receiver, we introduced a shutter controlling algorithm to enable multiple access in VLC without compromising higher data rates capability which eventually improved the overall system performance. Then, the thesis focused to utilize VLC enabled active LED transmission in object detection and localization applications. Using optical blinking sequence of LED transmitter, we proposed an optical correlation based signal decoding algorithm in camera receiver to localize the light emitter precisely which can be applicable in detecting and localizing identical objects. This correlation algorithm has the capability to be integrated in the existing visual SLAM (Simultaneous Localization And Mapping) techniques to enhance the system performance by improving the detection, localization, and tracking accuracy. In summary, this thesis addressed the following open research questions or challenges of VLC and mobile vehicular VLC applications through several novel systems and algorithms designing, and their prototype implementations:

- How to improve the signal to interference and noise ratio (SINR) and lower the bit error rate (BER) in VLC to enhance the system throughput, even in mobile environments?
- How to identify and then cancel or disallow the incoming noise & interference signals on VLC receiver?
- How to enable multiple access in VLC without compromising the higher data rates or system throughput?
- How can we implement visual tag features of VLC to localize and track identical objects (vehicles) precisely?

- How can active LED transmission enable precise and accurate localization without any prior information of LED transmitter's position or other related parameters?

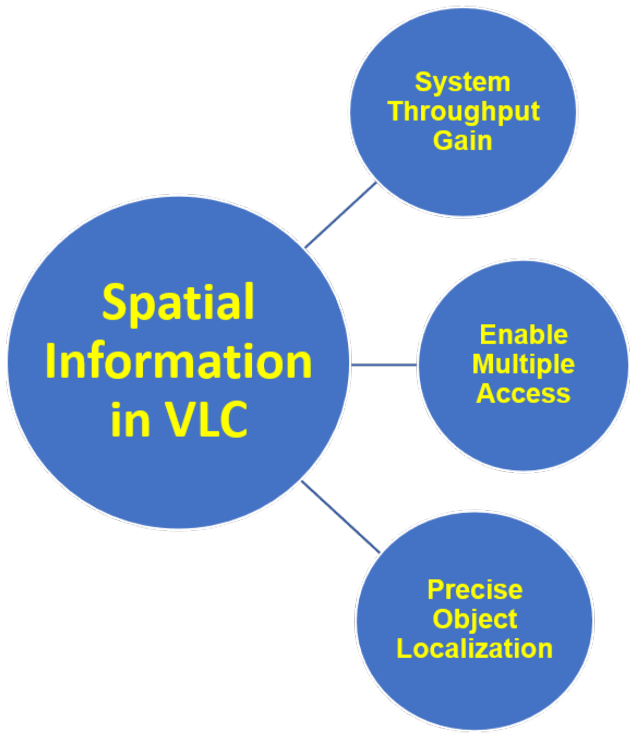


Figure 1.3: Thesis contributions: Spatial features in VLC systems.

1.3 Contributions

This thesis represents the following research contributions of mine throughout my doctoral journey to solve some open fundamental challenges in VLC and VLC enabled mobile applications, such as vehicular VLC:

Contributions 1: VLC System Throughput Gain by Spatial Filtering

- **Designing of a Novel Pixelated VLC Receiver using Spatial Optical Filtering**

This research approached the problem of achieving high system throughput in VLC from the perspective of a high-speed receiver design. In this part of the thesis, we presented a new VLC architecture to achieve high signal quality reception through a hybrid design that can leverage the advantage of photodiodes to achieve high data rates and the noise isolation property of image sensors. The hybrid design acts as single pixel ultra-high-speed-camera which has been validated through a proof-of-concept experimentation with significant SNR improvement [16; 17].

Contributions 2: Enabling Multiple Access in VLC through Spatial Resolvability

- **Spatial Multiplexing using Pixelated Shutter**

  In this work, we proposed, designed and evaluated a novel architecture for VLC that can enable multiple-access reception using a single photoreceptor receiver (photodiode). The novel design includes a liquid-crystal-display (LCD) based shutter system that has been automated to control and enable selective reception of light beams from multiple transmitters [16; 18].

- **Automated Shutter Control Protocol Design**

  To identify and separate noise and interference from the desired optical signals, this thesis introduced an automated shutter controlling algorithm (fast spatial tracking mechanism) in the pixelated shutter receiver. In our research efforts, we have demonstrated the feasibility of our VLC receiver architecture by conducting measurements study of noise and interference identification and separation using our designed shutter controlling algorithm [16; 18].

Contributions 3: Precise Object Localization through VLC Spatial Locality

- **VLC Embedded Optical Sequences Correlation to Localize Identical Objects**

  In this research effort, we developed a camera based visual identification solution using our proposed spatio-temporal optical correlation based localization algorithm. Trace-based evaluation of the identification or localization accuracy under real-world conditions including indoor, outdoor, static and mobile scenarios, showed that our designed optical correlation outperforms the comparative traditional machine learning (ML) and non-ML techniques for LED detection or localization [19].

CHAPTER 2

SPATIAL OPTICAL FILTERING BASED VLC RECEIVER PROTOTYPE DESIGN

Theoretical models estimate visible light communication (VLC) data capacity to be of the order of Tera-bits-per-second (Tbps). However, practical limitations in receiver designs have limited state-of-the-art VLC prototypes to (multiple) orders of magnitude lower data rates. In light of the technological challenges in VLC systems this research work introduces a new hybrid architecture to realize ultra high-speed visible light communication systems. The key idea of our proposed design is to leverage the fast sampling rates of photodiode receivers and integrate an image sensor–like shutter mechanism that filters noise and interference. Through adaptive selection of the exact receiver area over which the transmitted light is detected, the signal-to-noise-ratio (SNR) can be dramatically increased yet not compromising the high sampling rate achievable using state-of-the-art photoreceptors.

2.1 Introduction

The growing number of mobile devices and applications is straining the capacity of wireless mobile spectrum and has created what can be referred to as spectrum-crunch [20]. The significant growth in the wireless data traffic has initiated the need for expanding the range of frequencies used for wireless data communication. This has opened up new opportunities for utilizing the unused bands of the electromagnetic spectrum such as optical frequencies for wireless data communication through the Visible Light Communication (VLC) technology [21] [3] .

Today, VLC is going through an interesting transition from a purely academic concept to standardization through the efforts of IEEE 802.15.7 task group [22], and to commercialization through the concept of Li-Fi [23]. However, even with such rapid advancements in the technology, the state-of-the art data rates in VLC is dramatically less than its actual wireless data capacity. Achieving ultra-high data rates in VLC close to its wireless data capacity is the key vision of this proposal.

Considering the insufficiency of bandwidths in today's wireless technologies, *achieving ultra–high–speed VLC is not only an opportunity but also is a necessity.* Achieving data rates close to capacity in VLC requires significant advancements in science and engineering of highly efficient and robust VLC architectures. The fundamental issue with traditional VLC architectures is that, photodiodes can sample light signals at extremely high rates but signal quality suffers under high ambient noise scenarios. Multiple–input Multiple–Output (MIMO) through photodiode arrays and imaging receivers can spatially isolate noisy pixels due to the definite array structure, however, are extremely limited in sampling rates. Such architectural differences create a data–rate versus signal quality trade off in VLC.

2.2 Related Works

The survey paper [24] presents a consolidated list of existing VLC systems and the challenges in the domain from a scientific research perspective. The survey paper [25] discusses those challenges from a standardization and commercialization perspective. We will review state–of–the–art developments in achieving high speed VLC. In addition, we will also review some of the new application dimensions in VLC to give an idea as to how the technology is diversifying as a promising wireless technology.

LiFi. The state–of–the–art in commercialized VLC technology is the LiFi–X system [23]. LiFi–X includes a modified LED light bulb transmitter and a receiver hardware dongle with USB support that connects to a PC. LiFi–X is capable of 40Mbps uplink and downlink duplex VLC using a white LED transmitter and a high power, high–cost avalanche photodiode receiver. However, even adding an extra photodiode in this receiver can be extremely challenging due to the form–factor limitations, driving amplifier load, and the firmware overhead for processing an additional receiver element.

IEEE 802.15.7 standard. While the IEEE VLC standard theoretically supports data rates upto 96 Mbps [22], the simulation studies in the draft revision to the standard, IEEE 802.15.7r1 [26], claim Gbps data rates capability using orthogonal frequency division multiplexing (OFDM) [27] modulation. OFDM requires knowledge of the channel parameters to allot sub–channels for multiplexing, and this is proposed to be achieved by channel estimation using feedback loops. The practical viability and reliability of such designs can be extremely challenging considering scale and mobility.

Multiple–Input Multiple–Output (MIMO). The concept of MIMO, using arrays of LEDs and photoreceptors, has gained prominence in VLC architecture design. Using array transmitters and receivers allows for scaling the the data rate by the multiplexing data communication across multiple LED–photoreceptor channels. Multiple array elements also increases the field–of–view of the receiver thus allowing for some mobility within LOS. Array photoreceptors can be either a set of photodiodes arranged in a specific fashion or correspond to a set of pixel elements of an image sensor. The challenge with photodiode arrays is they allow more *noise*, from ambient light (sunlight and artificial lighting), into the receiver due to the wide field–of–view, thus affecting received signal quality and data rate. To account for this it requires very high amplification and noise reduction which can be complex and costly [28]. Image sensors can help isolate the noise because of its spatial structure, however, are extremely limited in sampling rates or frame–rates. Even the fastest image sensors can sample only at the order of 1000 FPS [29], which is orders of magnitude less than that of a single photodiode ($10^6 - 10^9$ samples/second).

Free–space optics inspired. Recent work in fiber–wireless–fiber based architectures in free–space optics design, estimate data rates of the order of 10s to 100s of Gbps [30]. These systems use high power and high cost elements such as Laser diodes controlled by optical–fiber elements at the transmitter/receiver. These systems require bulky spatial light modulators (SLM) to direct the laser beam using mechanical steering to cover a wide angular range, if not, use high–cost avalanche photodiodes at the receiver [28]. Due to the high cost, high power and complex hardware design, such architectures will not be appropriate for generic VLC systems.

The Smart Lighting Research Center [31] at Boston University identifies the design of fast–switching and power–efficient LEDs for smart space solutions as one of its key research thrusts. Other thrusts include, LED–to–LED communication [32], power–line VLC networks [33], duplex VLC [34], backscatter VLC [35]. These efforts promote the diverse

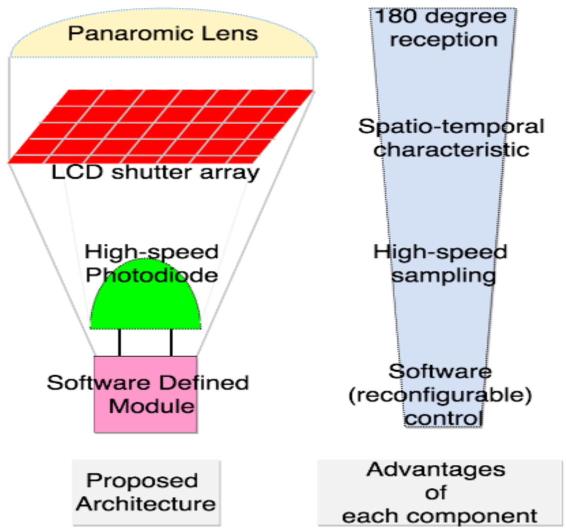


Figure 2.1: Proposed pixelated shutter based hybrid VLC architecture

use–cases of VLC. However, there is a consensus that high data rate VLC system design is a need of the day.

2.3 System Design

We combine the advantages of photodiodes and image sensors, and propose a novel receiver architecture that *emulates the functionality of image sensor arrays using a single photodiode.* The core idea of this design is to utilize the high–speed sampling of a photodetector and augment features of a typical image sensing array. We provide a conceptual overview of the architecture in Figure 2.1.

The key components of the proposed design are the high–speed photodiode, a shutter mechanism, the computing unit and a panoramic lens. The shutter mechanism enables to spatially filter the noise and interference from the actual optical signal from the light source.

In this regard, we use an off–the–shelf liquid–crystal–device (LCD) shutter array[36], where each element of the LCD array, or *pixel*, doubles up as a digital shutter based on the input voltage. Depending on the input voltage, the liquid crystals occupy a certain polarity thus allowing light to traverse through the pixel only if the polarity matches that of the incoming light beam, if not blocks the same. The receiver uses this functionality to control which light beam must be processed and what must be eliminated by the photodiode. The computing unit enables high sampling rate processing and hosting a software stack to incorporate control and other processing mechanisms. A typical software defined radio (SDR) unit, such as an Universal Software Radio Peripheral (USRP) [37] or an FPGA device, can serve as the computing unit. A panoramic lens fit to the PD–LCD array will provide a wide–angle (180 degree) FOV to the receiver. The LCD array with the lens expands the effective FOV of the photodiode yet preserving its high sampling rate and eliminates the need for multiple photodiodes to achieve the array structure. The LCD array and photodiode will be controlled independently using the computing unit. Such a modulo hardware architecture makes this design re-configurable.

2.4 Implementation and Evaluation

The strength of the proposed hybrid architecture lies on two fundamental notions, that, (a) light sampling can be controlled using a digital shuttering mechanism, and (b) unwanted optical signals can be eliminated by separating signal from noise and interference directly in spatial domain.

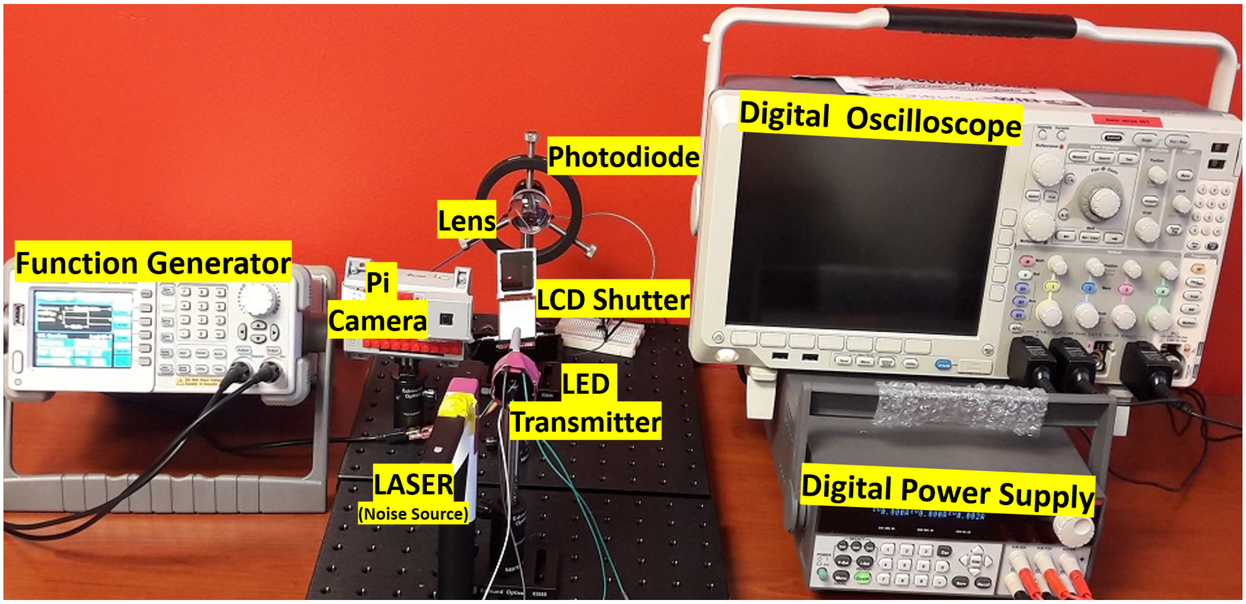


Figure 2.2: General experiment setup

Through a proof-of-concept experimentation, in our initial research effort [16], we have studied the feasibility of noise and interference reduction through our hybrid architecture.To this end, we meticulously arranged an optical measurement setup on an optical table to carefully quantify the signal, noise and interference signals. We ensured there are no vibrations or any movement that can impact the quality of our measurements. The measurements were conducted indoor, in an academic lab. The lighting involved the ceiling florescent white lamps and the ambient sunlight from across the room through the glass window. The experiment was conducted on the less bright side of the room at a distance of 20ft from the glass window.

The general experiment setup is shown in Figure 2.2, and consists of an off–the–shelf PIN photodiode [38], a red LED [39], a laser LED (acting as noise source) [40], an TFT LCD shutter [41]. We used a Keithley 2231A-30-3 digital power supply [42] to power our

LEDs and a Tektronix digital oscilloscope [43] to monitor the output from the photodiode. We also used a digital multimeter to record the photodiode output voltage and current. We setup a RaspberryPi camera [44] on a plane parallel but translated from the LCD shutter center for visual verification.

To ensure the photodiode signals are registered on the multimeter and the scope, we amplified the photodiode output using a LM358N operational amplifier [45]. We used the circuit in non–inverting mode with a resistance of $R = 510k\Omega$ and the voltage output $V_o = RI_{pd}$, where I_{pd} was the received photocurrent. In the setup, we used an off–the–shelf aspheric condenser lens [46] to focus the light wave onto the photodiode. The focal length of the lens is 27mm and the photodiode was placed at the focal point of the lens in our experiment. The lens was placed behind the shutter covering the area of the shutter. The distance between the lens (shutter) and the photodiode is 2.7cm (focal length of the lens) and the distance between the shutter and LED transmitter is 16cm.

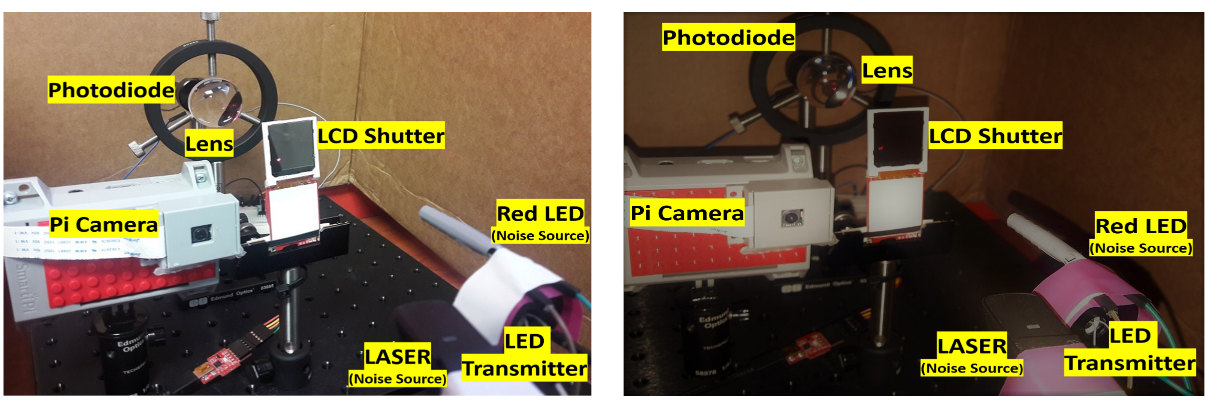


Figure 2.3: Our measurement setups for (left) Open box testing and (right) Closed box testing

Figure 2.3 shows a closeup view of our experiment setups. We setup two modes for our experiments, (a) closed box, where we covered the setup using a cardboard box to create a dark-room type environment by blocking the ambient light, and (b) open box, where we let the top part of the box open while one of the sides was covered with cardboard to block the sunlight from the glass window. In our experiments we physically blocked the sunlight and hence the ambient noise in our experiments is primarily from the ceiling white lights. As you can observe from the setup figures, we used a LASER LED and another RED LED light source which played the role of interfering (noise) sources for the primarily LED-Photodiode link.

2.4.1 Spatial Noise Filtering

Using the setup shown in Figure 2.3 (b), we conducted an experiment to measure the SNR for different choices of reception area on the LCD shutter. We conducted the experiment in a closed–box setting to ensure no ambient lighting impacted the noise measurement. Hence, the noise measured in this experiment corresponds to the limited ambient lighting within the box (negligible) and the noise from the signal (typically very low). The main goal of this experiment is to understand the relative SNR improvements if the area of the reception was centered around the area of the photodiode.

The experiment involved shining an LED in direct current mode (no modulation) on the photodiode by concentrating through the center pixels of the shutter. The LED transmitted 1mW of optical power. We ensured that that the angle between the transmission and reception axis was zero. We measured the received voltage on the photodiode with the LED in OFF mode, as V_n, and as V_r when the LED was ON. We compute the signal–to–noise–ratio

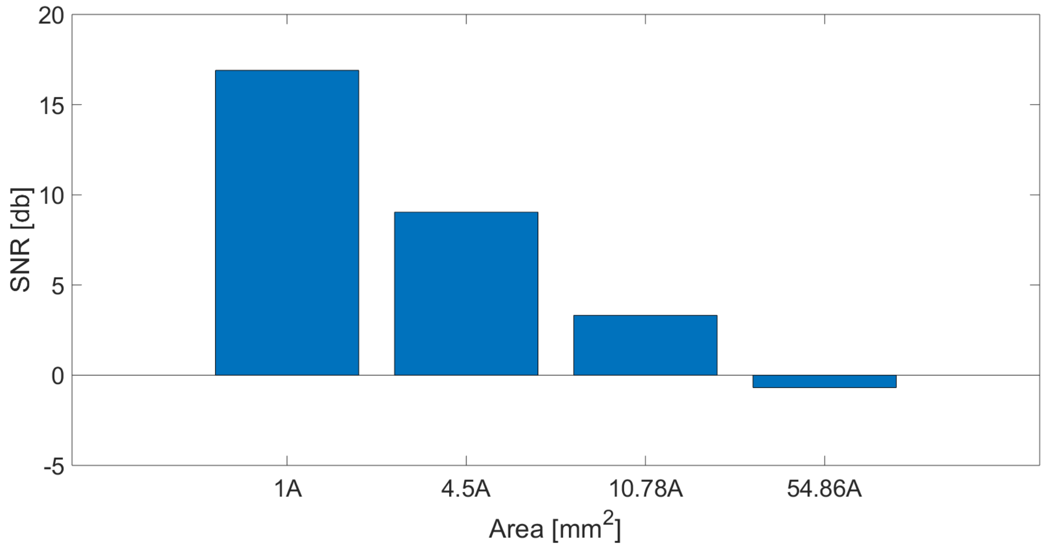


Figure 2.4: SNR versus selected area of reception. Here A = 57.40mm^2. Set resolution of LCD shutter = 240 x 320 pixels, side length of the LCD shutter pixel = 0.2mm

as,

$$SNR = \frac{V_r^2 - V_n^2}{V_n^2} \;\; \& \;\; SNR_{dB} = 10\log_{10}(SNR) \tag{2.1}$$

We conduct the measurements for different area of shutter opening. For each area of selection we open the appropriate number of pixels considering it as a square region. Consider A = 57.40mm^2, we conducted these measurements for 4 area selections:

(i) 1A: Only the area corresponding to the actual area of the photodiode was open

(ii) 10.78A :An area in between the LED illumination and PD surface area was open

(iii) 4.5A: Only the area corresponding to the LED illumination on the shutter was open

(iv) 54.85A: Entire shutter was open

We plot the SNR versus different areas of shutter opening in Figure 2.4. We report the minimum SNR over 10 trials for each area setting. We can observe from the SNR decreases significantly with increase in shutter opening area. This is in line with our theoretical understanding of the dependency of SNR on area of reception. In particular, if the receiver has a larger area of opening, it allows for more photons to be registered on the photodiode. However, if the desired signal occupies only a fraction of that area, the rest corresponds to accumulating noise and other undesired photons. Due to the additive nature of photon energy, separating signal from noise becomes extremely challenging if the SNR is low. The SNR values suggest that, if it can be ensured that the receiver photodiode is collecting only the photons corresponding to the actual signal, then the effect of noise on the receiver becomes almost negligible. The improvement in SNR, as can be observed from these measurements, is such that when the area corresponding to the exact photodiode area is opened while other parts of the shutter are closed, the SNR is at 17dB, compared to the -1dB SNR value when the entire shutter was open. The negative SNR indicates that the noise component over powered the signal and hence demodulation is impossible. We also observe that the area of reception corresponding to the LED (4.5A) is not necessarily the best choice. This is because, the LED signal when projected on the shutter, acts like a diffuse source. The optical energy from the LED is distributed over a larger area (than the photodiode), thus, relatively, allowing for more noise photons to be registered at the receiver. It is also notable that 18dB increase in SNR can be considered dramatic in terms of communication systems and are usually achieved only through extremely sophisticated and complex signal processing. Our measurements suggest that it may be achieved through a rethinking of the receiver hardware.

2.4.2 Spatial Interference Cancellation

We conducted an interference and noise cancellation measurement experiment using the open box and closed box setups shown in Figure 2.3. We treat the RED LED as light source, another RED LED as interfering source and the RED LASER as noise source. The ambient lighting in the room (white light from ceiling) is accounted as an additional noise source. We conducted measurements of received signal voltage V_r, noise voltage V_n and interference voltage VI across open box and closed box setups, and along two modes of area selection: (i) Shutter fully open, (ii) Only pixels corresponding to noise LED and interference LASER LED projections were closed. During these measurements, all the light sources were input with a constant power and were set to operate in their maximum optical power output (supply maximum forward current). We compute the SINR using the voltage measurements as,

$$SNR = \frac{V_r^2 - (V_I^2 + V_n^2)}{V_I^2 + V_n^2} \tag{2.2}$$

Setup	V_r	V_I	V_n	SNR_{db}
Openbox+Shutter open	5.0	3.7	0.7	-1.18
Openbox+Shutter (I+N) closed	4.1	0	0.7	15.22
Closed box+Shutter open	4.9	3.6	0.4	-0.81
Closed box +Shutter (I+N) closed	4.2	0	0.2	26.43

Table 2.1: SINR measurements in open box and closed box setups when all optical sources are in always-ON (DC) mode. All voltage values are in Volts. Shutter (I+N) closed means the pixels corresponding to interference and noise projections on the shutter were closed. The closed box setup was not a totally dark setup. There is slight ambient light entry which was measured and calibrated to be 0.2V.

We report our measurements and SINR values in Table 2.1. We can observe from Table

2.1 that when the shutter is fully open or when the interference area was closed, the signal power almost remained constant. However, the interference can significantly overpower the signal if it were allowed to register on the photodiode. We can observe that the laser source which has a significantly higher optical power than the LEDs can completely overpower the system and hence lead to SINRs that are almost useless (close to zero or negatives). We also can observe that, even with an overpowering interfering source, through spatial filtering, the SINR can be dramatically improved, in ranges of 15-25dB. In the next set of experiments, we set up to modulate the LEDs using a single frequency pulse waveform. We connected the signal and noise LEDs to the GPIO pins of RaspberryPi .We modulated each LED using a separate Raspberry Pi, which was controlled using MATLAB on a laptop. The waveform input to the LED was generated in MATLAB and communicated to the LED via the RaspberryPi. The signal LED was modulated at 300Hz (the pulse waveform read as 295Hz due to some distortions in the RaspberryPi link) with a 8V peak-peak pulse waveform. The noise LED was input with a 3V peak-peak pulse waveform at 100Hz. The laser LED was set in DC mode.

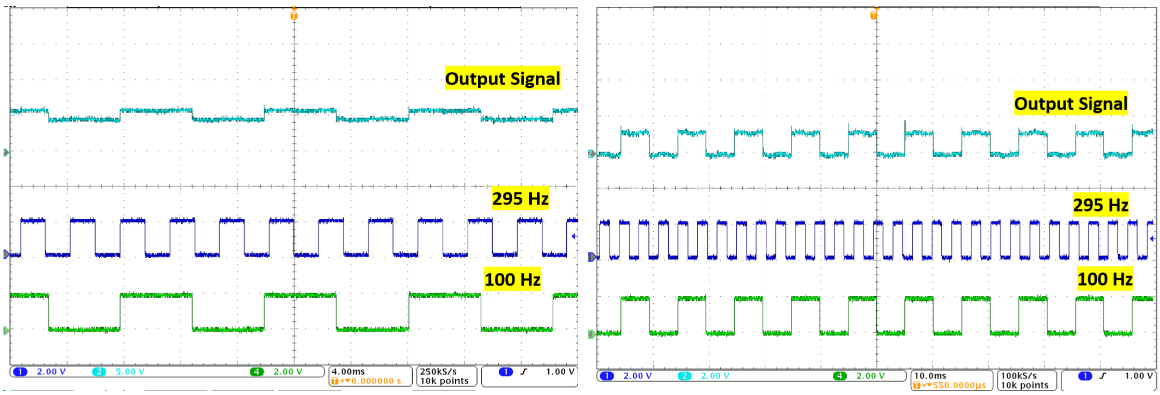


Figure 2.5: Illustrating additive interference from alternating (AC) signals on the photodiode receiver. The measurement was taken with the shutter opened and shutter closed in closed box setup.

We can observe the additive property of optical signals in Figure 2.5 . We can observe that the two signals plus the DC noise (laser beam) is added in the output. When signals of the same frequency are accumulated on the phototodiode, due to a phase difference of 0 (or 2π), the resultant signal is essentially an amplified version of the original signals. When the phase difference is non–zero (or not2π), then the effective phase-shift will be captured in the additive signal on the photodiode output. Assuming, we know at least one of the transmit frequencies, through a cross correlation mechanism we can find out the phase difference and hence differentiate the wave forms at the receiver. However, the temporal separation of the waveform will be possible only when the receiver can ensure that it is exactly sampling the signals and not any unwanted optical energy. Also it requires knowledge of at least one of the frequencies in the set. Hence, resolving the signal from noise from this cumulative signal is extremely challenging without calibrating the noise and interference levels, which adds complexity and usability constraints on the VLC system.

2.5 Conclusion

This work presents a new architecture that combines the high-speed sampling advantage of photodiodes with the spatial filtering capability of image sensor receivers. We presented a hybrid architecture design that uses a high-speed photodetector and an LCD shutter acting as a programmable image sensor aperture. As the first step, we conducted measurements to study noise and interference separability in our receiver. Our measurements indicate that the spatial separability, if achieved correctly, can help improve the signal quality in the receiver and almost completely eliminate noise. In this preliminary research outcome, the notion of the prototype implementation was to show a proof-of-concept understanding. Through the knowledge gained from the measurements using this setup, in future, we will design a custom receiver that leverages the advantages claimed by the design.

CHAPTER 3

SPATIAL MULTIPLEXING ENABLED MULTIPLE ACCESS USING A SINGLE PHOTODIODE VLC RECEIVER

The directionality of optical signals provides an opportunity for efficient space reuse of optical links in visible light communication (VLC). Space reuse in VLC can enable multiple-access communication from multiple light emitting transmitters. Traditional VLC system design using photo-receptors requires at least one receiving photodetector element for each light emitter, thus constraining VLC to always require a light-emitter to light-receptor element pair. In this paper, we propose, design and evaluate a novel architecture for VLC that can enable multiple-access reception using a photoreceptor receiver that uses only a single photodiode. The novel design includes a liquid-crystal-display (LCD) based shutter system that can be automated to control and enable selective reception of light beams from multiple transmitters. We evaluate the feasibility of multiple access on a single photodiode from two light emitting diode (LED) transmitters and the performance of the communication link using bit-error-rate (BER) and packet-error-rate (PER) metrics. Our experiment and trace based evaluation through proof-of-concept implementation reveals the feasibility of multiple LED reception on a single photodiode. We further evaluate the system in controlled mobile settings to verify the adaptability of the receiver when the LED transmitter changes position.

3.1 Introduction

Visible Light Communication (VLC), is an emerging wireless communication technology that operates unregulated in the visible–light band (400–800 THz frequencies or 380–780 nm wavelengths) of the electromagnetic spectrum, and is enabled by light emitting elements such as light emitting diodes (LED) and light receiving elements such as photodiodes (PD). Due to directionality of light beams, VLC is a line-of-sight (LOS) technology that requires the light transmitter and receiver to be within each others field-of-view (FOV) [47]. The LOS requirement provides novel opportunities for efficient space and time reuse in VLC where multiple light emitting transmissions could be multiplexed.

Traditional VLC [2] that operates using a single non-array photodiode receiver based reception, requires to incorporate specific multiple access mechanisms to enable reception from different light emitters. By leveraging the directionality of optical signals and that light emitters can be spatially differentiated, it is possible to multiplex signals by combining space division multiple access (SDMA) with time/frequency/code (TDMA/FDMA/CDMA) division access schemes, however, the nature of photoreceptors to collectively add all the detected photons within its FOV limits makes differentiation of multiple transmissions and from ambient noise very challenging. This limits VLC to the effective communication using only one light beam (or transmitter) at each instance of time. The key challenge in using multiple access mechanisms in single-photodiode non-array VLC receiver systems is that the incoming signals, through may be spatially and temporally separated, but once they reach the receiver collector (lens), the signals are mixed (leading to interference) with each other and thus cannot be differentiated. Unlike radio-frequency communication, where polarity of signals and thus representing signals as complex numbers is possible, in optical wireless, the received signals are essentially the positive-sum of all photons, which carry no polarity.

With stringent constraints that there cannot be an extra channel and that the signals must be easily identified from each LED on the photodiode, multiple access in VLC is very challenging without the use of array receivers or side information. Therefore, unless there is extra information regarding the signals (possibly through an extra control channel) or custom detection mechanisms incorporated into the system, the ability to spatially and temporally differentiate optical signals using a single non-array photodiode receiver remains an open challenge.

Multiple-Input Multiple-Output (MIMO) architectures for VLC that have been proposed and designed before [9; 28] require multiple photoreceptor elements, for example, as in photodetector arrays and image sensing arrays or cameras. The use of array elements in a receiver limits the sampling bandwidth (photodiode sampling frequency or camera frames-per-second) of the receiver hence limiting the achievable throughput of the system.Therefore, enabling multiple-access while retaining the high-speed sampling capacity of photodiodes is the other key open challenge that remains to be addressed.

To address the challenges presented above, this paper explores the use of spatial filtering mechanisms using a new hardware design for VLC receivers, to enable multiple access. In essence, this work presents a proof-of-concept study of using liquid crystal displays (LCD) to potentially enable multiple access in single non-array photodiode receivers. We propose a receiver design that uses LCDs to differentiate multiple LED transmissions and enable multiplexed communication using a single photodiode in the receiver. To this effect, we build our system over a baseline architecture from our prior work [16], where a liquid crystal display (LCD) panel was used as a digital gate or *shutter*, to allow or disallow signals onto a photodiode receiver. In this paper, we extended this baseline design to build a single photodiode receiver equipped with a LCD shutter array controlled by an automated signal selection protocol to differentiate reception from multiple LED transmissions (conceptual diagram in Figure 3.1 (left)). This protocol helps relax the assumption (considered in our prior work) of the apriori knowledge of which LCD shutter should be opened (or closed), and enables on-the-fly determination of the intended shutter state. The proposed receiver architecture, to be referred to as *pixelated shutter* receiver for the rest of the paper, sets the foundation for future VLC system architectures to achieve MIMO communication using only a single photodiode receiver. Achieving MIMO first requires multiple access reception capability demonstration, and to be best of our knowledge, our work in this paper presents the first design and evaluation of a novel multiple access VLC receiver. In summary, the key contributions of this paper are:

1. Design of an automated shutter control protocol for selective reception of multiple LEDs using a pixelated shutter receiver.

2. Implementation of a prototype *pixelated shutter* receiver multiple access system with 2×2 LCD panels, single photodiode and a software-defined radio.

3. Experimental trace based evaluation of the *pixelated shutter* receiver which employs the bespoke proposed automated shutter protocol, for 2 independent LED transmitter scenario using (a) bit-error, (b) packet error and (c) latency metrics.

4. Experimental evaluation of the automated shutter control protocol under controlled mobile environment.

3.2 Related Works

Background. In our prior work [16], we introduced a new architecture for VLC that uses a high–speed photodetector and an LCD shutter acting as a programmable image sensor aperture. In this work, our measurement studies proved that noise and interference can be separated spatially using our VLC receiver to improve the Signal to Noise Ratio (SNR) and Signal to Noise Interference Ratio (SINR) significantly. Through a proof-of-concept experimentation, in our previous work [16] we have studied the feasibility of noise and interference reduction by manually selecting one of the shutter pixel apertures for higher signal reception. In this paper, we relax the assumption (considered in our prior work) of the a priori knowledge of which shutter should be opened (or closed), and advance the design by proposing a novel automated shutter control to help differentiate LED signals on the photodiode. This paper leads and consolidates the idea of multiple access in VLC by adopting a shutter controlling algorithm in the receiver. We evaluate the system for high speed reception (up to 2MHz data transmission frequency) for multiple access from 2 LEDs and feasibility of automation under controlled mobile settings.

In the rest of this section, we discuss some of the existing works that are closely related to the challenges targeted in our system. However, we emphasize that no prior work has shown the capability of achieving high-speed multiple access in VLC using a single photodiode receiver, which remains the key focus of this paper's contribution.

Existing VLC technologies. Recent years have seen active development of VLC research prototype systems. However, these systems have specific limitations owing to the customizations in their hardware/software design. PureLiFi [23] devices have shown the capability of Mbps-Gbps data rates, but their performance is limited to static VLC settings and carry high hardware overheads, complex signal processing, thus requiring a huge cost for the product. OpenVLC1.3 [48] is an open-source embedded VLC prototype based on simpler protocols that offers data rates upto 400kbps. The customized hardware and requirement of modifying the operating system kernel makes this option very challenging to generalize. With a 1.4kbps data rate, LocalVLC [49] presents a low-cost VLC prototype for indoor IoT applications based on morse code modulation, however, fails to motivate its usage for typical indoor IoT applications where higher throughput is typically required. Other VLC systems including Purple VLC [50; 51; 52], have similar challenges as discussed above. With achieving high data rate remaining a key challenge for existing VLC systems, the addition of multiple access requirement can either require significant modifications to these designs or may not be feasible at all. Our proposed design can potentially address high-data rate, multiple access and mobility all at once through a unified design. By using the shutter control mechanism efficiently, signals can be differentiated from noise and interference, thus enabling cleaner signal reception or higher Signal-to-Noise (SNR) and/or Signal-to-Interference-and-Noise Ratios (SINR). With a cleaner signal, the signal modulation mechanisms can be kept simple with more focus laid on achieving multiple access, high-throughput reception and mobility.

Non-Orthogonal Multiple Access in VLC. Recent works have proposed non-orthogonal multiple access (NOMA) schemes for VLC [53; 54; 55] to improve spectral efficiency and enable multiple access in VLC. However, the common challenge in these designs is the reliance on channel state information (CSI) and the requirement for the transmitter and receiver to be informed a priori about CSI. For example, the work in Reference [53] introduces a power domain based multiple access protocol so that the users can use the entire bandwidth during the communication session, but requires the CSI and only works in small indoor environments. NOMA is a good contender for multiple access in VLC however the designs have been largely limited to showing the feasibility of interference cancellation under strong assumptions which limit the effective throughput performance of the VLC system.

Multiple Access using MIMO Techniques. Using the MIMO technique proposed in Reference [56], several works have presented different multiple access schemes using different equalization [57] and modulation schemes such as OFDM [58; 59], optical spatial modulation (OSM) with OFDM [60]. Reference [61], introduces an Optical Code Division Multiple Access (OCDMA) technique and Reference [62] used intensity modulation to support multiple users in MIMO VLC system. A key challenge with photodiode arrays is that they allow more noise, from ambient light (sunlight and artificial lighting), into the receiver due to the wide field–of–view, thus affecting received signal quality and data rate. The work in Reference [63] proposes multi-color LEDs based MIMO VLC system to ensure higher data rates (upto 1Gbps) but the system requires complex signal processing and equalization techniques. Reference [64] presents a precoding technique to mitigate inter-cell and intra-cell ambient light interferences in multi-cell VLC systems to improve the bandwidth efficiency, where specific spatial regions are considered as cells, similar to cellular communication. The complexity of the MIMO techniques for real-time implementation and performance, and the necessity of high efficiency and costly photoreceptors (avalanche photodiodes [28]) for improving data rates, remain challenges yet to be solved for MIMO VLC systems.

Wavelength Division Multiplexing (WDM). In Reference [65], the authors introduce a bi-directional VLC in full duplex mode by parallel transmission of three (RGB) channels and use OFDM modulation demodulation to increase the aggregate data rate. In another work [66], a color-shift keying CDMA (CSK-CDMA) based VLC system has been developed to increase the VLC throughput and for allowing multiple access.To provide ultra high data rates ($>$ 35 Gb/s) in a wide range of coverage, a WDM system of four-colour multiplexed using MEMS based beam-steering has been presented by Chun et al. [67]. While such systems could potentially avoid interference across specific wavelengths, the complexity, high bit-error-rates, and costly hardware elements limit the usage of these approaches.

3.3 System Design

We have designed a pixelated shutter based VLC receiver that automatically identifies and selects/isolates signals from multiple LEDs. The key components of the system, as illustrated in Figure 3.1 (left), include a photodiode (capable of high-speed sampling), a LCD shutter array, a shutter control unit, a computing unit, and a condenser lens for optical focusing. The key idea of proposed design is to allow signals from multiple LED transmitters to be correctly detected and decoded using a single photodetector VLC receiver. With the knowledge and assurance that only one LED signal impinges on the photodiode at each of its sampling instance, the receiver can be operated at the bandwidth matched to the LED's transmission, resulting in high SNR and thus potentially high data rate reception.

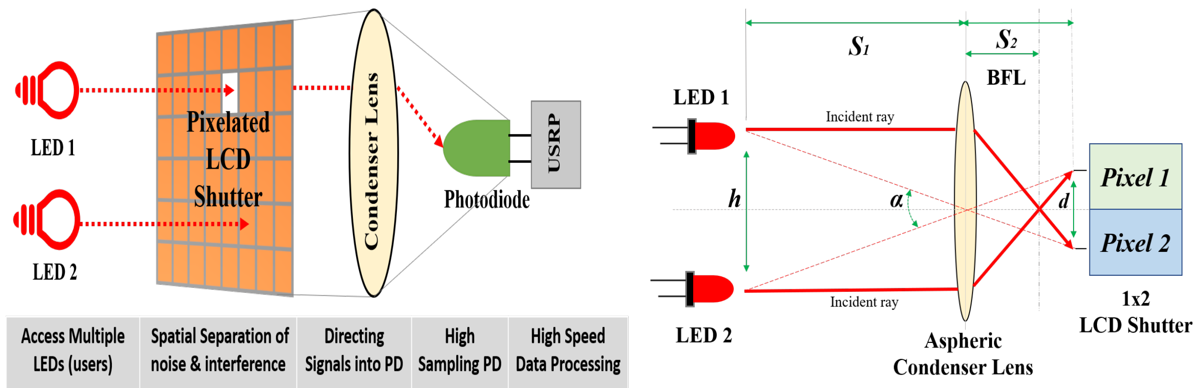


Figure 3.1: (**left**) Conceptual diagram of *pixelated shutter* visible light communication (VLC) receiver. (**right**) LED Positioning geometrical analysis to ensure focus of only 1 LED per shutter pixel.

The shutter uses LCDs which act as a digital aperture that allows (disallows) the impinging light beams, to reach the photodiode, based on the input voltage to the shutter. Using this digital aperture as a control the receiver is able to select which of the incoming light beams are to be decoded by the photodiode at each instance of time. The computing unit at the receiver hosts the decoding algorithms and mechanisms to efficiently decode the signal that has been selected. The digital control of the shutter is integrated with the decoding modules in the computing unit, such that there is active feedback on the quality of the received signal. The feedback information includes the received signal-to-noise-ratio (SNR) and a digital identification of the signal using packet header bits. This design enables a seamless functioning of the selective control of the reception and the decoding in tandem. The selection of the desired signal(s) is a one-time process and needs to be repeated only when there is mobility or during link failures.

3.3.1 Spatial Multiplexing using Pixelated Shutter

We describe the multiple access capability through the spatial multiplexing setup illustrated in Figure 3.1 (right). Consider two LEDs placed in space (at same height but separated along the horizontal) aiming to focus their signals (using the condenser lens) onto a photodetector (not shown in Figure 3.1 (right)) by passing through a 1×2 LCD shutter pixel system, where the pixels are aligned next to each other along the horizontal axis. Let us consider that each pixel i is responsible for signals from corresponding LED i. In this way, when the signals from the LEDs are beamed onto the photodiode, each pixel can selectively allow/disallow the signals provided the signals are independently identified (and differentiated) and the information on which signal (LED) should be selected is feedback to the shutter control unit (not shown in Figure 3.1 (right)).

To ensure that the signals do not overlap onto a single shutter surface area, our design requires that each LED signal can be spatially separated onto independent shutter pixels. This depends on the size of the shutter, distance of communication and the spatial separation of the LEDs. Through lens equation [68] and using simple trigonometrical calculations, we derive that the minimum distance of separation between the LEDs must be $h = dS_1/BFL$ where d is the distance between two shutter pixel centers (considering a square pixel it is the pixel side length), S_1 is the distance from LED to the lens, and BFL is the back focal length which is the focal length of the lens. We also derive the minimum angle of separation between the LED beams as $\alpha = 2\arctan(\frac{h}{2S_1})$. These equations provide the designer the control of placing the LEDs in space so as to allow multiple access reception on the single photodiode receiver. We determine that the minimum horizontal separation between the LEDs has to be $h = 14.88cm$ and $\alpha = 51.2^o$ for the component values from our proof-of-concept prototype system (section 3.4), where $S_1 = 15.5cm$, $S_2 = 8.2cm$ and $BFL = 3.75cm$ and $d = 3.6cm$, radiation angle of the LED is 50^o. at a distance of 10m the equivalent minimum separation distance would be close to 10m, however, this distance can be reduced if the d were to be increased. By merely doubling the size (d) of the shutter, the required distance h now can be 5m. This means that the selection of separation between LEDs in space and the size of the shutter involves a tradeoff. The tradeoff between h and d can also be adjusted by using LASER type emitters or LEDs which have smaller radiation angle.

3.3.2 Automated Shutter Control Protocol Design

The shutter control protocol operates in tandem with the decoding process running in the computing unit connected to the photodiode. The control protocol, discussed in Algorithm 1 and demonstrated as block diagram in Figure 3.2 involves two steps:

Step 1: Discovery phase, where the receiver does a preliminary pruning of all signals that do not represent a transmit signal by using the signal-to-noise-ratio from each shutter pixel i (SNR_{pxi}). SNR is computed as ratio of signal power to noise power, where power is computed as the mean-squared photodiode voltage reading divided by the sampling time interval. This step helps to filter ambient light, DC noise sources and other known noise sources. This way, only a subset of the shutter pixel array are kept OPEN and are to be processed, thus limiting the processing to a smaller subset of signals.

Step 2: Identification phase, where the receiver does a fine tuning of identifying each transmit signal and selectively opening the corresponding shutter pixels to allow/disallow the signal for continued reception at the photodiode. The identity of the signals are maintained through unique header sequences (barker codes) in the data packets. We consider that a unique ID of each transmitter will be registered at the receiver apriori during the first setup of the system (one time) and update the identity look-up table as necessary.

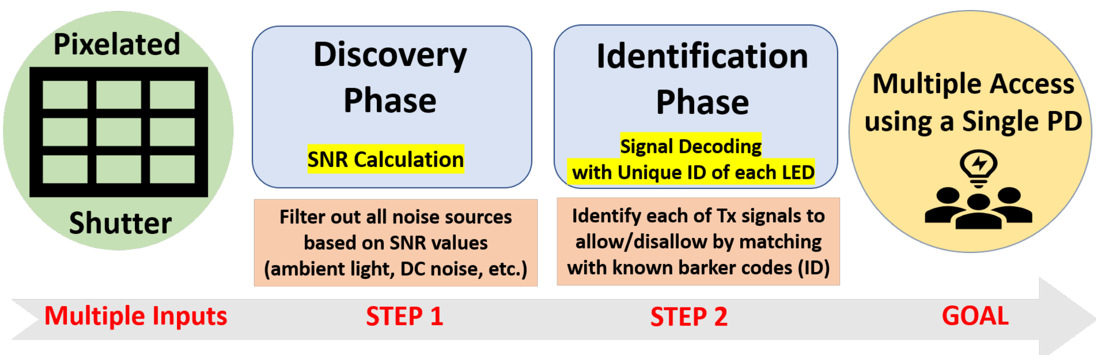


Figure 3.2: Block diagram of shutter controlling algorithm.

The shutter control protocol enables multiple access where information from multiple LED transmitters can be decoded by a single photodiode receiver. By enabling which signals to receive at which instance of time, the receiver can choose time-slots to receive and decode specific signals. Our system by default functions as a space and time-division multiple access (SDMA and TDMA) system where each transmission is decoded across a time-slot duration of T_s seconds, and the selection of spatially separated LED emissions is controlled through the LCD functions. While transmitters could potentially transmit at different rates and that time slots of transmission and reception may incur synchronization issues, the spatial separation through the LCD enables to first find which signal is intended and which is not, and then physically allow only the signal intended. Without the LCD the other possibility to achieve the same functionality is to use an array receiver, however, as mentioned before, array receivers can significantly increase the complexity of the system. This work aims to study how a non-array receiver could still function such multiple access schemes, albeit with some practical and minimalist hardware additions.

The choice of the value of T_s depends on the application. For example, a slot duration of 1–2 seconds may work for beaconing and repetitive transmission such as in sensor or IoT applications, however, for streaming applications the slot has to be made much smaller (order of few ms). A smaller slot duration also implies that the pixel switching control must happen as fast as the selected slot duration. Depending on the type of LCD shutters, the switching time can vary from few micro to 10s of milli seconds.

3.4 Prototype Implementation

We implemented a prototype pixelated shutter receiver as shown in the setup in Figure 3.3. The design parameters are all listed in Table 3.1. The key components of the hardware system include 2 RED LEDs, a PDA10A2 Amplified Photodetector, a custom made 1×2 pixelated LCD shutter and an aspheric condenser lens (outer diameter 80 mm and BFL = 37.5 mm). We implemented the automated shutter control algorithm in a Raspberry Pi 3 Model B+ which interfaces with a 2×2 pixelated shutter, built using off-the-shelf LCD shutter elements from AdafruitDue to the leakage of LED light we use the 2×2 LCD setup in 1×2 reception mode. We used two N210 USRPs as the computing units

Algorithm 1 Automated Shutter Control Protocol

OUTPUT: Allow desired optical signals and block interference

Initialization:
Open all shutter pixels ($px1$, $px2$, $px3$.....pxn) to allow signals
Set a fixed switching time (T_s) for all pxi ($i = 1, 2...N$)
Preset an empirical threshold SNR value (SNR_{th})
Refresh unique identifier (ID) look-up table

Step 1: Discovery
Iterate each pixel OPEN and all others CLOSED for duration T_s and record SNR
IF $SNR_{pxi} \geq SNR_{th}$
(a) OPEN shutter pixels pxi
(b) CLOSE all other pixels
(c) Proceed to **Step 2:Identification**
ELSE CLOSE all the pixels and refresh the program

Step 2: Identification
Correlate decoded signal ID with look-up table IDs
IF ID matches
(a) OPEN only matched ID containing pixels
(b) CLOSE all the remaining pixels
ELSE CLOSE all the pixels and goTo **Step 1: Discovery**

at the transmitter (controlling LED transmissions) and receiver ends (decoding signals from photodiode). We used a LFTX daughterboard capable of operation from 0–30MHz and a RFTX daughterboard. The 2 LEDs were controlled using two different USRPs, each hosting a LFTX, and one of the USRPs hosting a LFRX that also conducted the reception. We used GNU Radio blocks (block diagram of GNU applications shown in Figure 3.4) to transmit and receive signals using the USRPs. We chose to use the state-of-the-art Gaussian Minimum Shift Keying (GMSK) as the modulation strategy in our design, however, any type of modulation can be used in the system. We use 13-bit and 11-bit barker sequences for LED1 and LED2 header bits, respectively. We have implemented the transmissions in the form of UDP packets of size 2096 bits (as per IEEE 802.15.7 VLC standard [69] packet definitions).

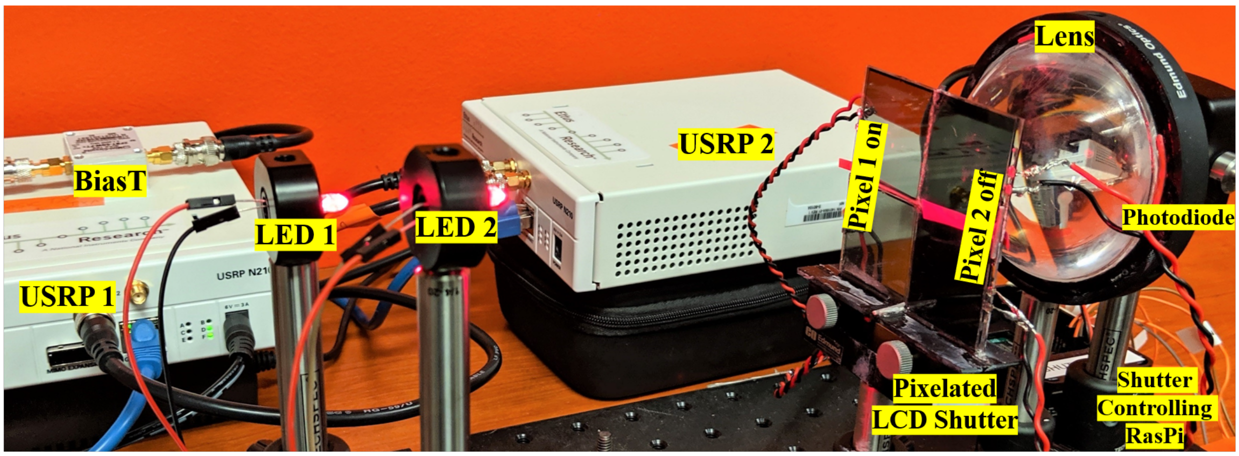


Figure 3.3: General setup of the pixelated shutter receiver system. This picture shows a 2 LED transmitter setup with a single photodiode receiver and 2 × 2 liquid-crystal-displays (LCDs) fit in a 1 × 2 shutter pixels configuration.

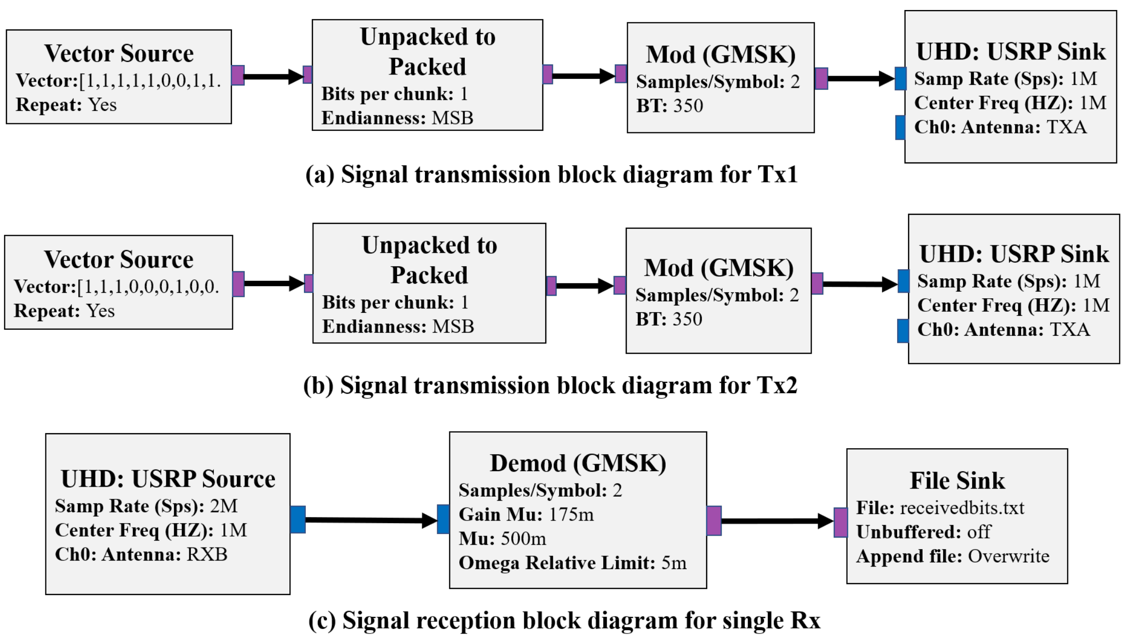


Figure 3.4: USRP GNU Radio block diagram for (**a**) transmitter LED 1, (**b**) transmitter LED 2, a and (**c**) pixelated shutter receiver. Note that the vector source values shown are only example values. The setup uses 2 N210 USRPs with LFTX daughterboard for the transmitters and 0-30MHz LFRX daughterboards for the receiver.

LED Specifications	
Number of LEDs	02
LED Type and Size	Red, T-1 3/4 (5 mm)
Horizontal distance between LED's [cm]	14.88 cm
Viewing Angle of LED [deg.]	30°
Luminous Intensity at each LED [mcd]	3500 mcd
Optical Output Power at each LED [mW]	125 mW
Wavelength of each LED [nm]	650 nm
Photodiode (PD) Specifications	
Physical Active Area of the PD [mm^2]	0.8 mm^2
Wavelength Range of PD [nm]	200 to 1100 nm
Bandwidth of PD [MHz]	150 MHz
Peak Response of PD [A/W]	0.44 A/W
Lens Specifications	
Lens Type	Aspheric Condenser Lens
Outer Diameter [mm]	80 mm
Back Focal [mm]	37.5 mm
LCD Shutter Specifications	
LCD Type	TN, Transmissive, Positive
Dimensions of LCD Pixel	36 × 36 mm
Driving Voltage of LCD [V]	3–5 V
Maximum Opaqueness (%)	95%

Table 3.1: LED, Photodiode, Lens and LCD specifications

3.5 Evaluation

We evaluate our system to study the feasibility of our system to achieve high speed reception and the performance of the automated shutter protocol for multi channel visible light signal reception on a single photodiode receiver. All the experiments were conducted in a lab setting, indoors, under ceiling white ambient lighting. Unless mentioned, the distance between the LED and photodiode in our experiments was set to 15.5 cm.

3.5.1 BER Analysis

We conducted two types of experiments to evaluate the bit error rate performance of our system: (a) BER under different types of interference, and (b) BER under selective mapping of signal and shutter pixels. We calculated the BER as the ratio of total number of bit decoding errors to the total number of transmitted bits per trial. The BER experiments involved, in general, transmitting a random stream of 30,000 bits and logging the decoded bits at the receiver. Each experiment trial was repeated 5 times, BER was computed per trial and the average BER is reported. Unless otherwise specified, the BER values reported in this paper refer to the average BER over 5 experiment trials. We chose LED 1 as the desired transmitter and LED 2 as interference. The LEDs were modulated using the baseband signal from the USRP, where a pulse waveform at a specific (generation) transmit frequency was input to the LEDs which mapped a 1 to pulse HIGH and 0 to pulse LOW. Since the operation was in baseband, the transmit frequency is essentially equal to the transmitter data rate. The data rates were thus chosen as per the experiment goals:

- Goal A: The 100 bits/sec is chosen arbitrarily, as the primarily goal of this experiment is to validate the additive nature of optical signal at the photodiode receiver.

- Goal B: We chose operation at (500k, 1M, 2M) bits/sec rates and measured the BER and PER at different shutter configurations at those rates for our proposed pixelated LCD architecture. The 2Mbps (2MHz) limit for evaluation was due to the limited operable range of the LFTX/LFRX USRP daughterboard. In our future work we intend to evaluate our system for Gbps range using SLD Laser type transmitters and using FPGA (Field-programmable Gate Array) computing nodes.

3.5.1.1 BER and Interference Patterns (Goal A)

We set the transmit frequency to be 100Hz and conducted the BER experiment under four signaling types:

- **Type 1:** Only the transmit signal.
- **Type 2:** The transmit signal and ambient DC noise.
- **Type 3:**Transmit signal and Interference signal sending identical patterns in phase.
- **Type 4:** Transmit signal and Interference signal sending identical patterns at 180 deg out-of-phase.

We report the BER from these experiments in Table 3.2. We can observe from the BER values from Case 1 that the error rates for the system are generally high when the desired signals and the interference are combined at the photodetector. We observe that the BER is practically low (for feasible data communication) when the pixels are selectively OPEN/CLOSE to allow only the desired signal, which has been achieved without major changes to the receiver. We can also observe from the BER values, the additive property of the receiver, where the BER is low when the interference signal is identical and of same phase (as the effective received signal amplitude is doubled) and high when the same interfering signal is out-of-phase.

Type 1		Type 2		Type 3		Type 4	
Case I	Case II	Case I	Case II	Case I	Case II	Case I	Case II
5×10^{-4}	5×10^{-4}	4.9×10^{-1}	8×10^{-4}	2.2×10^{-4}	2.1×10^{-4}	4.9×10^{-1}	5.1×10^{-4}

Table 3.2: Bit error rate (BER) at 100 Hz signaling under different interference patterns. Case 1: All pixels OPEN, Case 2: Only desired signal pixel is OPEN.

3.5.1.2 BER and Selective Signaling (Goal B)

Consider pixel-1 as the pixel corresponding to LED 1 and pixel-2 as the one for LED 2. We conducted the BER experiments under three different configurations of the shutter pixel and under three different transmit frequencies. The results from Table 3.3 indicate that the BER is at least an order low when only the desired signal is received versus when the interference is also sampled on the single photodiode receiver. The BER values, though relatively high (which can be reduced using error control coding) for data streaming applications, however, indicate the feasibility of multiple access using our proposed architecture. Considering $Goodput \approx (1 - BER) * transmitsymbols/sec * bits/symbol * errorcontrolcoderate$, we note with an assumed code rate of 1/2, 2 Mhz (2 M symbols/sec) transmit rate and 2bits/symbol modulation rate, the effective $Goodput$ per LED in our preliminary system is about 1.9 Mbps.

Frequency (Hz)	Configuration 1	Configuration 2	Configuration 3
500 KHz	0.015	0.039	0.21
1 MHz	0.015	0.035	0.21
2 MHz	0.015	0.030	0.21

Table 3.3: BER at different shutter pixel configurations. **Configuration 1:** ONLY pixel-1 is OPEN. **Configuration 2:** ONLY pixel-2 is OPEN. **Configuration 3:** BOTH, pixel-1 and pixel-2 are OPEN.

We observed after our experiments and analysis that the positioning of our custom built LCD array on pixel 2 location was slightly tilted thus causing a focussing issue of any light beam falling on it to the photodiode. The lens was placed at exactly the optical focal length distance from the photodiode to ensure convergence of the rays, however, due to the tilt the LED 2 signal falling on pixel 2 was actually defocused. After a breakdown of the equipment we measured that the signal intensity was reduced by almost 50%. This actually confirms our finding that the BER is little more than 2x that of LED1-pixel 1. We report these numbers as is and believe it is a honest representation of our experiments and that it actually helps make key observations.

3.5.2 PER and Signal Selection

We conducted 5 PER evaluation trials each for considering LED 1 or LED 2 as the desired signal. We collected traces from these trials and determined the PER through offline calculations. In each trial we transmit a continuous stream of packets of size 2096 bits, where each packet has a random stream of bits as payload and a 13 bits header. The header served as the unique ID for each LED. The header would be the same for all packets from a specific transmitter. For LED 1 we chose a 13 bits sampled from barker sequences as ID, and for LED 2 we chose a 11 bits sampled from barker sequence plus a 2 bit (11) padding. We chose the time slot (T_s) duration to be 2 seconds for each iteration over a pixel. We recorded the received and decoded bits from packets in each iteration of pixel OPEN cycles. We collected the received signal traces from each iteration of each STEP of the automated shutter control protocol (Algorithm 1). We iterated over 1 cycle of each pixel OPENING and then choosing either of the pixels that corresponds to the desired signal. We can observe from Table 3.4 that the PER is around 3%-6% which is comparable to typical PERs observed in traditional multiple access wireless communication systems. We estimate the theoretical throughput considering a PER range of 3% to 6%, and error control code rate of 1/2 as 0.94–0.97 Mbps, where $Throughput \approx (1 - PER) * packets/sec * bits/packet * errorcontrolcoderate$.

Freq(Hz)	**# Pkts in Ts**	**Packet Error Rate (PER) (%)**	
		Pixel 1	**Pixel 2**
500 KHz	477	5.88 %	5.46 %
1 MHz	954	4.83 %	2.63 %
2 MHz	1908	3.36 %	3.25 %

Table 3.4: Packet error rate (PER) calculated from received signal traces from each pixel OPEN duration of $Ts = 2$ sec. Pixel 1(2) corresponds to signal from LED 1(2).

3.5.3 Impact of switching latency

The switching latency is the effective time taken by the shutter receiver and its associated processing to switch the state and control from one pixel to another. Switching latency can impact the signal quality from each pixel during the shutter control phases. The switching latency is a function of the per-pixel *time slot duration* T_s and the intermediate state (open/close) *switching time per pixel*. We compute the switching latency= $[(pixels - to - scan) * (T_s + switching - time - per - pixel)] + [(no. - of - transmitters * packet - size * switching - time - per - pixel)]$, where the first part of the sum is the Step 1 latency and second part corresponds to Step 2. We first validated the consistency of the SNR values under different time slot duration selections for the pixels in our current prototype. Considering LED 1 as desired signal and with LED 2 switched OFF, we alternated pixels 1 and 2 to be OPEN for the specific time slot duration and recorded signal and noise power. We report the average SNR values in Table 3.5 and observe the consistency of SNR for short (100 ms) as well as long time slot durations (2 s). The experiments were conducted in a well lit (white ceiling lighting) lab environment. The ambient light from the ceiling light is considered as the noise source, with a recorded average voltage was 4.9 mV across all experiments; to help compare, the LED signal from 16 cm recorded about 0.6 V on the photodiode without any shutter.

Pixels	**Pixel 1**	**Pixel 2**	T_s **[ms]**
SNR_{db}	19.97	−0.27	100 ms
	19.96	0.48	500 ms
	19.86	−0.47	1000 ms
	20.02	−0.60	1500 ms
	19.99	−1.18	2000 ms

Table 3.5: Average Signal to Noise Ratio (SNR) in dB for each pixel OPEN under different shutter switching times. Here, the LED transmitter is placed such that it illuminates Pixel 1 only.

In our prototype the time slot duration is 100ms, pixel switching time is 1ms, and there are 2 transmitters sending 2096 bits packets. We measure the Step 1 (discovery) phase switching latency to be about 400 ms ($2 \times (100$ ms $+ 1$ ms$)$) with a 5 s periodicity, and Step 2 (identification) phase switching latency to be about 4.2 s ($2 \times 2096 \times 1$ ms). Ideally, the smallest shutter pixel switching time is desired, however, the hardware choice may cause a constraint. We observe that our LCD shutters in our prototype can go lowest up to 1ms operation. If we were to consider practical usage of switching times of $T_s = 1$ microsecond (potentially using Digital Micromirror Devices DMDs), and shutter resolutions of 100×100 and 1000×1000 pixels, and even considering an overestimated number of 100 effective transmitters (which map to 100 different pixels) the effective, theoretically estimated, processing time (latency) of Steps 1 and 2 in Algorithm 1 would be about 220ms and 1.2 s, respectively. Such latency numbers can be considered practical for typical VLC applications including sensing, IoT and low-speed device-device data transfers.

3.6 Extended Evaluation: Mobile Scenario

We extend the evaluation of our prototype system across mobile environments. The mobility considerations in these evaluations refer to the the case when the VLC transmitter can potentially change its spatial position while transmitting data. The receiver is kept stationary in all these experiments. When an actively transmitting LED changes its position, the receiver must be able to actively identify that the movement has happened and that it needs to adapt its reception area. Using our pixelated shutter approach, we hypothesize that when the LED changes position, the receiver will identify the movement event based on its periodic SNR measurements across each pixel and the automated shutter control protocol

(Algorithm 1) shifts control to the appropriate pixel over which the LED signal is now being received. The goal of these evaluations is to verify the feasibility of our system to adapt its reception when the LED changes position.

3.6.1 Experiment Setup and Methodology

To facilitate controlled movements of the LED transmitter across different positions, we set up our VLC system on a table top with the receiver in static position. We 3D printed a housing for each of the two LED transmitters to be integrated with a 3-wheel RaspberryPi controlled robot as shown in Figure 3.5. With a distance of 0.5 m set between the VLC transmitter and receiver, we consider three key positions of the transmitter; A, B and OFOV (out of Field-of-View):

1. Position A is the point from where the LED transmitter illuminates Pixel 1 area of the shutter. When the robot is in position A, the receiver must be able to identify the signal is on Pixel 1 and only OPEN Pixel 1 (keeping others closed).

2. Position B is the point from where the LED transmitter illuminates Pixel 2 area of the shutter. When the robot is in position B, the receiver must be able to identify the signal is on Pixel 2 and only OPEN Pixel 2 (keeping others closed).

3. Out of Field of Views (OFOV) are the marked positions on the straight line trajectory of the robot where the LED signals are either minimal or out of field-of-view of the receiver lens and photodiode. When the robot is in this position, the receiver must

identify there is no active transmission from the LED being received and hence must close all pixels to avoid any noise or interference signals being sampled on the receiver.

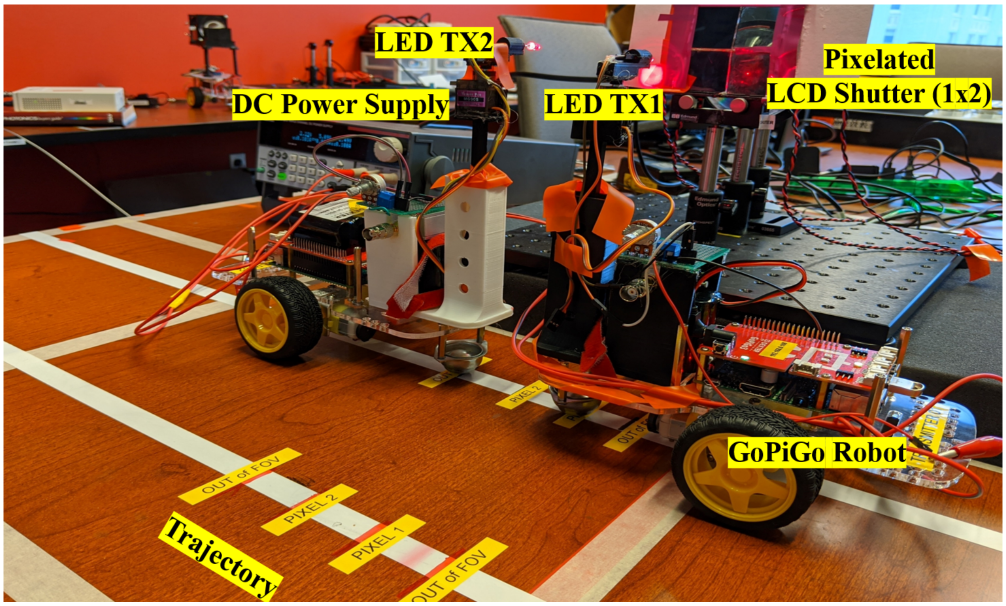


Figure 3.5: LED transmitters placed on GoPiGo RaspberryPi robots

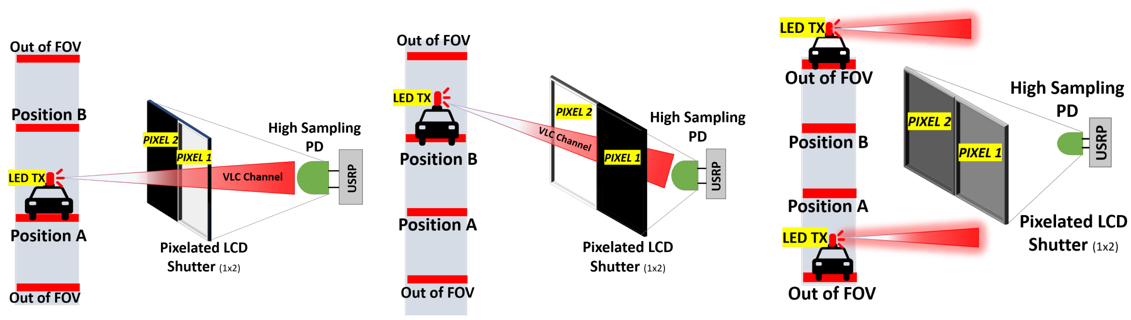


Figure 3.6: Mobility experiments with the variation of LED Transmitter's position: A (**left**), B (**center**) and Out of Field of Views (OFOV) (**right**).

3.6.2 Results

We set the robot to move from one end of the table to the other along a straight line trajectory, traversing positions A, B and two marked out-of-FOV points (see Figure 3.6 for an illustration). In each experiment trial we chose a specific speed of the robot and shutter switching time and move the robot along the calibrated trajectory. The receiver selects the pixel status as OPEN/CLOSE based on, first computing the SNR on each pixel and flagging the pixels which have SNR above an empirical threshold (Table 3.5) for each switching time selection, and then the receiver making a recommendation on which pixels to OPEN/CLOSE based on whether it has detected the ID sequence (decode bits, correlate with sequence and flag success if correlation above 90%). We set our experimentation to record the system's pixel shutter recommendations in the form of a 2 bit binary representation as listed in Table 3.6. For example, a 01 implies that the system correctly identified that Pixel 1 should be open and identified the signal on Pixel 1 using the 5 bit Barker ID sequence transmitted as payload.

Configurations	**Binary Representation**
Pixel 2 Closed, Pixel 1 Closed	00
Pixel 2 Closed, Pixel 1 Open	01
Pixel 2 Open, Pixel 1 Closed	10
Pixel 2 Open, Pixel 1 Open	11

Table 3.6: Binary mapping of the output or shutter pixels status recommendations

In Figure 3.7, we report our system recommendation output for three speeds of the robot (20 dps or 0.53 cm/s, 15 dps or 0.4 cm/s, 10 dps or 0.26 cm/s) and 1000 ms shutter switching time. There is an ERROR case of recommendation where both pixels are open (binary representation: 11) in both 15 and 10 dps robot's speed experimentation. During this time, the robot's position is at the intersection of A & B, where both pixels are illuminated at the same time and the measured SNR values in both pixels are higher than the threshold value. However, we observed over 5 trials, even in the reported least accurate case, the ERROR cases occur largely in the transition phase from one pixel to another. This can be resolved by synchronizing the transition time with the speed of the robot or by speeding up shutter sampling (smaller intervals) so that even if the error occurs during a transition, there are enough samples to identify the behavior and a suitable action can be taken on the receiver. Further, more receiver samples imply possibility of using probabilistic estimation of whether the LED has changed position or not.

Figure 3.7: Our system's output (recommendations) at 1s shutter switching time and at different speeds of the robot: 20 dps (**left**), 15 dps (**center**), 10 dps (**right**). We conducted 5 trials and ordered the highest to least accuracy. These figures represent the least accuracy case.

As an extended experiment we evaluate how the system performs for much smaller sampling time intervals such as 100 ms and 50 ms. These sampling intervals can cater to real-time human-computer interaction applications and other applications using real-time feedback. While the focus of this work is not necessarily to emphasize only one category of application, we do explore on the success and limitations of our system under such fast sampling time settings. We report the results from our robot movement based experiment results, discussed previously, for these cases and report them in Figures 3.8 and 3.9, respectively. We observe that in most sampled points the system recommends correctly, however, some cases of error when the LEDs are not in FOV or transitioning. These errors can be resolved by employing statistical filtering to the recommendation/identification results in each sample over the duration of a reasonable and practical time window. For example, even if the sampling duration is 100ms, the samples across a 1sec time window can be considered for statistical estimation and filtering noise to ensure the correct receiver sampling area is accounted. Another issue, the leakage of LED signals through the shutter even if it is CLOSED. We note that we have used an off-the-shelf LCD shutter, which has only a 95% marked opacity, which means that 5% of the ambient and incoming light on the shutter are still let through. At these short distances we believe the leakage is causing significant changes in the SNR values and hence leading to erroneous recommendations when both pixels are CLOSED or both OPEN. One option is to replace the LCD shutter with Digital Micromirror Device (DMD) based high speed shutters to ensure much faster and efficient pixel switching. In summary, we note that the goal of our current work is to explore these mobile scenarios and make observations of the artifacts arising in our system response. We reserve the optimization of the system performance for better hardware and across mobile scenarios for future work.

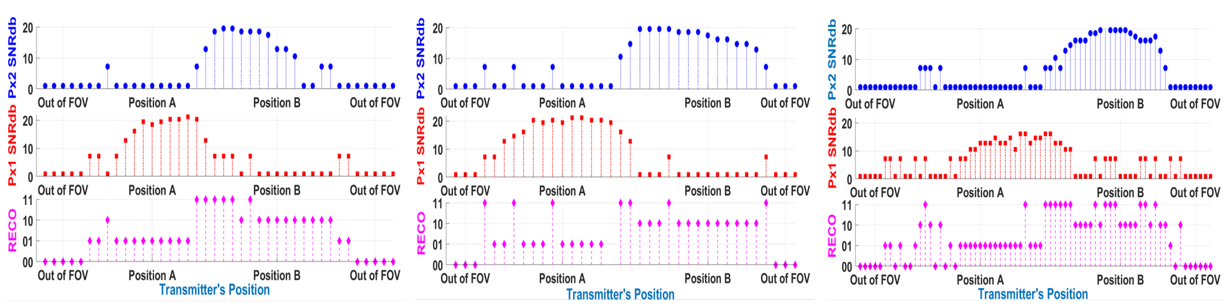


Figure 3.8: Our system's output (recommendations) at 100 ms shutter switching time and at different speeds of the robot: 20 dps (**left**), 15 dps (**center**), 10 dps (**right**)

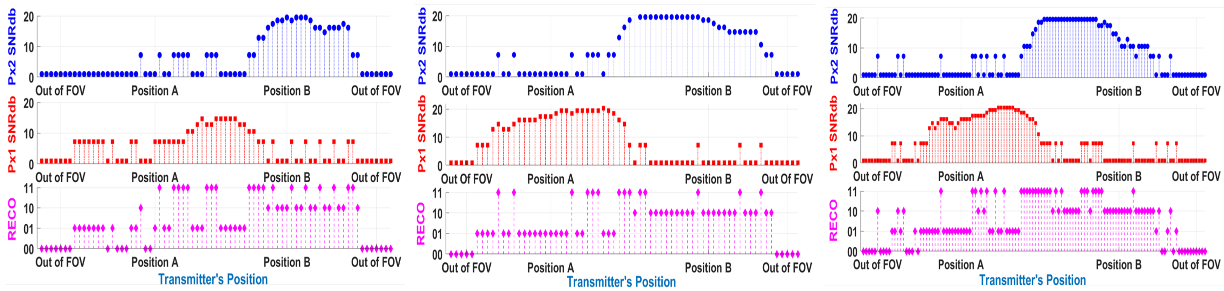


Figure 3.9: Our system's output (recommendations) at 50 ms shutter switching time and at different speeds of the robot: 20 dps (**left**), 15 dps (**center**), 10 dps (**right**)

3.7 Discussion

Ideally, the size of the LCD pixel area has to be matched with the size of the LED signal blob on the receiver. The size of the blob changes (decreases) with (increasing) distance between the transmitter and receiver. It will be appropriate to have an extremely small LCD pixel area to ensure the signal from the maximum range of the LED such that the signal on the photodiode is above ambient noise (usually meters to 10s of meters). In that regard, we originally chose to use a TFT LCD pixel array of a small TFT LCD monitor screen (used for Arduinos and Raspberry Pis). However, due to the hardware and software limitations in controlling separate pixels, we chose to build a contraption of an LCD pixel array and the ones we have used are the smallest size that can be obtained off-the-shelf. In our next iteration of this design we plan to use Digital Micromirror Devices (DMD) to achieve small areas of reception as well as faster switching times.

3.8 Conclusion

In this paper, we introduced a novel architecture and a protocol to enable multiple access reception on a VLC receiver with only a single photodiode element. We designed and evaluated a system that enables transmission from 2 LEDs simultaneously and selectively decodes packets from each based on a selection algorithm that uses OPEN/CLOSE cycles of LCD shutter pixels acting as digital apertures for the photodiode signal. Through BER, PER and latency metrics (processing latency), computed through experiments, we showed the feasibility and performance of our 2 transmitter-to-1 receiver multiple access system at low and high signal frequencies. We also successfully verified the feasibility of the use of the system across controlled mobile settings by setting the LED transmitters on robots and the pixelated receiver being static. To the best of our knowledge, this work sets the foundation stage for future work in multiple access using single photodiode receiver.

CHAPTER 4

CAMERA BASED LIGHT EMITTER LOCALIZATION AND TRACKING USING OPTICAL BLINKING SEQUENCES

Visual identification of objects using cameras requires precise detection, localization, and recognition of the objects in the field-of-view. The visual identification problem is very challenging when the objects look identical and features between distinct objects are indistinguishable, even with state-of-the-art computer vision techniques. The problem becomes significantly more challenging when the objects themselves do not carry rich geometric and photometric features, for example, in visual identification and tracking of light emitting diodes (LED) for visible light communication (VLC) applications. In this paper, we present a camera based visual identification solution where objects or regions of interest are tagged with an actively transmitting LED. Motivated by the concept of pilot symbols, typically used for synchronization and channel estimation in radio communication systems, the LED actively transmits unique pilot symbols which are detected by the camera across a series of image frames using our proposed spatio-temporal correlation based algorithm. We setup the visual identification as a problem of localization of the LED on the camera image, which involves identifying the (*pixels*) and the *unique ID* corresponding to the LED. In this paper, we present the algorithm and trace-based evaluation of the identification accuracy under real-world conditions including indoor, outdoor, static and mobile scenarios. In addition to micro-benchmarking the localization accuracy of our technique across different parameter configurations, we show that our technique outperforms comparative techniques, including, color based detection, support-vector machine based (SVM) machine learning, and you only

Figure 4.1: Depiction of different LED localization application scenarios: (left) indoor robot localization application, (right) outdoor 3D mapping, V2V and pedestrian localization using LED and camera.

look once (YOLO), which is a state-of-the-art convolutional neural network (CNN) deep learning based object identification tool.

4.1 Introduction

The advent of camera-based automation in mobile systems, advances in autonomous robotic systems and pervasive use of visual perception as an essential modality in cyber-physical systems, have urged the need for visual identification of objects in a given scene with high accuracy and precision. Fundamentally, this problem has long been studied and addressed along the dimensions of object detection/recognition and localization using computer vision. The advancements in deep learning have improved vision based recognition fidelity. Localization, along with 3D environment mapping, have improved significantly using visual SLAM (Simultaneous Localization and Mapping) [70; 71]; computer vision used with SLAM to build a map of an unknown environment and perform localization to locate the object or robot (self) inside the generated map.

Vision based techniques fundamentally reach a bottleneck when the objects of interest are identical, making differentiating objects using visual features alone impossible, and that when the environment is dynamic and mobile, thus causing problems for matching features across time for reliable visual SLAM. For example, an autonomous driving vehicle mapping the 3D environment suffers from distinguishing different identically looking buildings and other road side objects. The constantly changing scenery, due to motion, further complicates the process as the visual features are 'available' only for a short duration (even shorter depending on the speed of the vehicle). To address this issue, we propose that such objects in the scene, particularly those which can lead to such vision bottlenecks, be tagged with a light emitting diode (LED) which constantly transmits a unique ID (mapped to the object of interest in the scene) and a camera is used to *localize* this LED. The unique ID serves as a differentiator between objects, and the localization problem boils down to precisely identifying the pixels in the camera images that correspond to the LED. To this end, we propose a novel *correlation localization* technique that is fundamentally motivated by the concept of pilot symbols correlation used in radio packet communication reception. The pilot information in the form of barker code binary sequences are transmitted by the LED that are detected, demodulated on camera image pixels, and the corresponding sequence of digital data is cross correlated with the known pilot (barker code) sequence. A high correlation will mean that the particular camera image pixels correspond to the fact that the LED was detected at those pixels.

Correlation Localization. We setup the visual identification as a problem of localization of the LED on the camera image, which involves identifying the (*pixels*) corresponding to the LED. We treat that the unique ID for each LED in vicinity is registered in the camera system's database. Note that the purpose of these unique IDs is to differentiate the objects of interest within the scene in immediate vicinity of the camera. Thus these IDs can be reused and the number of IDs within a spatial region is finite and will scale linearly (with number of tagged objects of interest). Motivated by the concept of pilot symbols, typically used for synchronization and channel estimation in radio communication systems, the LED actively transmits unique IDs, or pilot symbols, which are detected by the camera across a series of image frames using our proposed spatio-temporal correlation based algorithm. This algorithm takes a window of image frames, registers the scene using compute vision image alignment technique, and performs a one-dimensional n-block correlation across the image – treating the image matrix of pixel intensity as a linear array of numbers. The n is the parameter that represents the number of elements in the array used for correlation. The fundamental idea is that only the pixels corresponding to the LED will follow a intensity variation pattern in accordance with the pilot symbols, while the other background pixels do not change significantly or are mostly static. This way, the pixels corresponding to the LED alone will reveal a high correlation output which thus helps isolate the LED pixel region with high accuracy and precision.

Applications. LED localization can be very helpful in a plethora of applications, particularly those relying on location based services and those which use cameras. As depicted in Figure 4.1 (left), LED localization can significantly assist in autonomous robot navigation and scene mapping. Active transmissions using LEDs and decoding using cameras is the fundamental concept of visible light communication (VLC). Hence, localizing a LED in itself fundamentally solves the key issue of transmitter identification and tracking in VLC. The concept of visible light positioning (VLP) has gained much interest in the research community for localizing ground objects based on locating LEDs and identifying them by decoding bits from LED transmissions. However, VLP depends on prior knowledge of the map or blueprint of LED placements and fundamentally tries to solve the dual problem (localize the camera device with respect to the local space based on detected LED positions using geometrical analysis). Accurate localization of the LED in the camera image will enhance VLP system fidelity. This is applicable even in outdoor scenarios (Figure 4.1 (right)) such as for mapping infrastructure (e.g. buildings), localizing safety critical events such as a pedestrian crossing the road, and for tracking target vehicle (transmitter and/or receiver) for vehicle-to-vehicle (V2V) communication (using VLC and/or radio wireless).

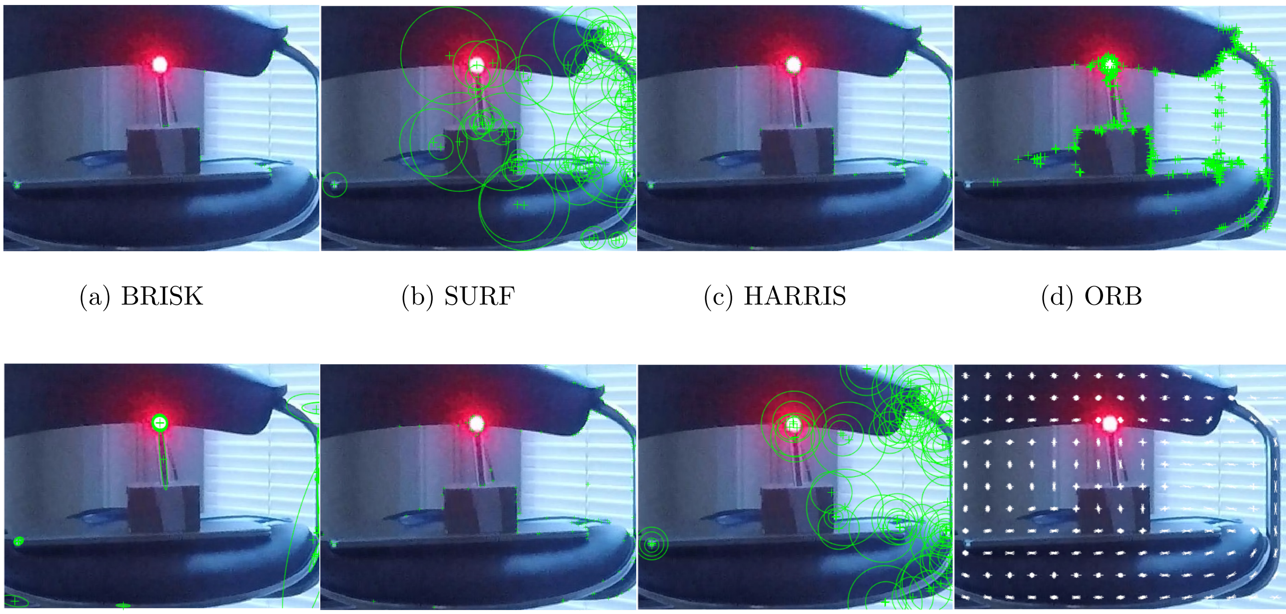
In summary, the key contributions of this work are as follows:

1. Design and implementation of the correlation localization algorithm for localizing LED on camera images.

2. Real-world trace based experimental evaluation of the correlation algorithm in different indoor, outdoor, static and motion (car driving) cases.

3. Performance comparison of the optical correlation decoding algorithm with color based thresholding and support-vector machine learning based localization accuracy metrics.

4. Comparative evaluation based discussion of advantages and disadvantages of using deep learning techniques for LED localization in camera images.



(a) BRISK (b) SURF (c) HARRIS (d) ORB

(e) MSER (f) EIGEN (g) KAZE (h) HOG

Figure 4.2: Various feature (key points) extraction techniques tested on an image of an LED switched ON.

4.2 Design Motivation: Challenges in Vision Feature Extraction

Features play a fundamental role in computer vision based algorithms; used for object detection, recognition, tracking, matching, classification applications and many more. Visual features in images, also referred to as *key points*, are essentially visual markers in the regions of interest (e.g. object) that can help characterize the particular image region. Computer vision algorithms for localization and tracking are fundamentally dependent on feature extraction from the scene, and every thing else that follows is largely based on the quality of the

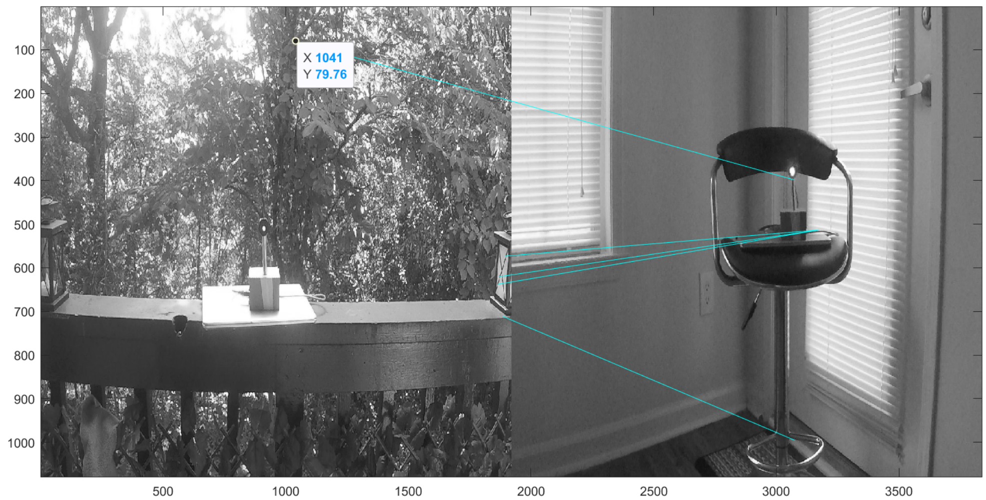


Figure 4.3: Implemented SIFT feature based feature (key points) matching for LED detection. Note that there are no SIFT key points on the LED region in both cases.

features. Some of the most prominent features that are used in computer vision applications include ORB [72], scale invariant feature transform (SIFT) [73], SURF [74], histogram of gradients (HOG) [75], Harris [76]. Other features that have gained prominence also include BRISK[77], MSER[78], EIGEN[79] and KAZE[80].

As a motivation experiment to demonstrate the challenge in LED localization using traditional vision techniques, we conducted a feasibility experiment with testing extraction of all the features listed in the previous paragraph. We used the MATLAB [81] computer vision toolbox to run each feature extractor on a sample image of a red color (monochrome) LED placed on a chair in room with some sunlight through the windows and no ambient artificial lighting. We can observe from Figure 4.2 that most of the feature extractors are not even able to find a single key point on the LED or close to the LED. Those that detect key points in this LED scene, such as SURF, KAZE and ORB, are very noisy as they are detecting multiple areas not representative of the LED as key points. Differentiating such key points is extremely challenging without much additional information, which is not the case. We can observe that HOG does detect some key points in a systematic manner, however, the problem of differentiating/cleaning the outliers is very challenging. In a more complex environment (backgrounds) the challenge will only become harder, as the key points will largely be concentrated on other aspects of the background that may have more visual characteristics than the LED. Clearly, the failure of traditional vision based feature extraction is attributed to the lack of knowledge or the ability to define features pertinent to the LED as it bears no clear and unique geometric or photometric characteristic.

As an additional measure, we tested the SIFT feature extraction and matching on the indoor LED scene, which worked better than the others, yet noisy. However, when the same LED was placed in a different setting – outdoor sunlight with trees background – the SIFT feature matching algorithm could not identify any credible key point on the LED in the outdoor setting and instead matches (wrongly) the indoor LED with the leaf region on one of the trees. This example is an additional evidence of the challenge in using feature extraction based techniques for LED localization.

4.3 Related Works

In this section, we survey related works on object detection and localization.

Feature extraction based Computer Vision. Conventional feature extraction based computer vision techniques using different descriptors such as SIFT [73], HOG [75], SURF[74], Haar[82] to detect and localize the objects from the scenes [83] are commonplace. Feature based extraction architectures are not robust enough to identify the objects accurately from the scenes due to the constant changes in the image backgrounds, illumination conditions and the appearances of the objects. LEDs in particular are feature-less objects making feature definitions for LEDs in real-world settings very challenging.

Visible light positioning (VLP). Using LED beacons can enable precise object localization through visible light positioning (VLP) [84]. Prior work has explored VLP across different applications such as, indoor localization, wearable devices, target tracking, etc [85; 86; 87; 88]. In VLP, the transmitter LED needs to send it's location information to the corresponding receiver (can be photodiodes or imaging sensors) to estimate the localization parameters including the distance and the direction of the light signals. However, such dependency of getting the information of position related parameters beforehand makes the VLP systems challenging especially in scenarios where the object's location and environment are unknown.

Learning based tracking and Re-identification. In intelligent transportation system (ITS), identifying, locating, and tracking the same or similar type of vehicles is still challenging for computer vision applications[89]. Recently, deep convolutional neural networks based approach has been extensively used to solve the vehicle re-identification problem in works such as PROVID framework [90], DRDL model[91], CityFlow [92], VeRi-Wild [93]. For example, in DEx [94], a CNN based dual embedding expansion technique was implemented to create unique representations from each of the images. However, all the techniques require large and diverse datasets of the object in question which can be a bottleneck.

Multi-sensor fused based object detection. Fusing information or data [95] from different sensors to detect and locate objects is one of the common research trends in the community for the last few years. Sensors data from different 3D detectors such as camera (both monocular and stereo)[96; 97; 98; 99; 100], LiDar [101; 102], Radar [103; 104] have been fused in several experiments to tackle the object detection problem. In [105; 106; 107; 108; 109], the authors propose different fusing techniques either by cascading the camera and LiDAR information or fusing the region of interest (ROI) features from the sensor information. In ContFuse [107], the system uses a convolution neural network based deep learning technique [110] to fuse ROI-wise the camera and LiDAR sensor data. To achieve full multi-sensor fusion, both point and ROI-wise features fusing have been implemented in [111]. However, fusing multi-sensor information is not an easy task to perform as there are challenges in every steps of data association, modality or alignment which needs a rigorous processing framework resulting in higher computational complexity.

4.4 System Design

The proposed system considers that objects or regions of interest in the space are tagged with a LED transmitter that serves as the meta identifier for the object and representative of where it is located within the scene. The LED is set to actively transmit unique IDs as a sequence of bits using on-off keying (OOK), where bit 1 is mapped to a high intensity level (ON status) and bit 0 is mapped to a low intensity level (OFF status) of the LED.

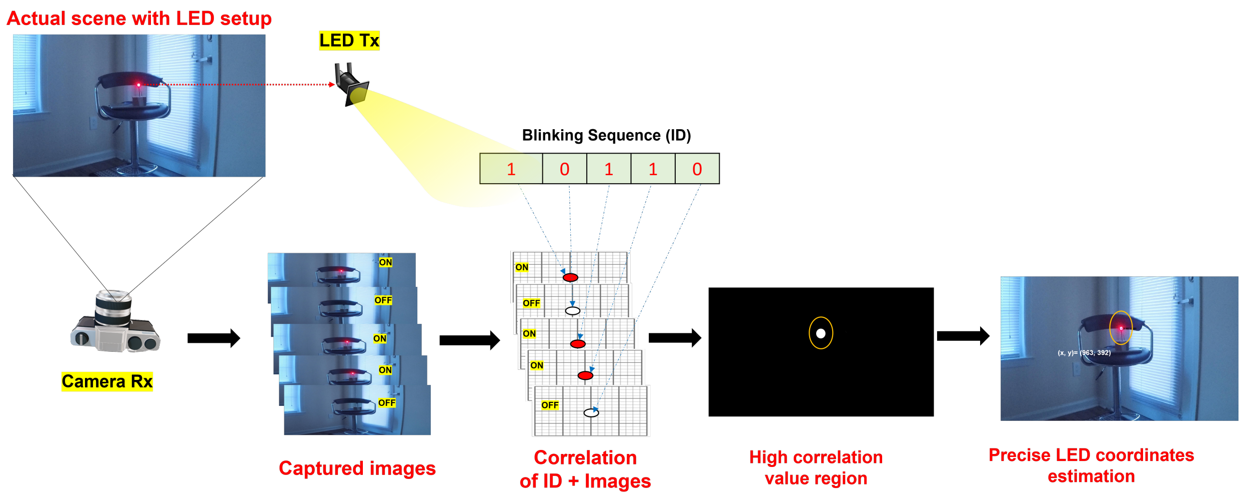


Figure 4.4: Design pipeline for proposed correlation based LED localization system.

We consider that each LED is set to a unique ID sequence, however, this sequence can be programmatically changed.

At the receiver, a camera that is perceiving the scene registers the LED signals; as long as the LED is within the camera's field of view, typically at narrow (± 30-50 deg) or wide (± 50-80 deg) angles for traditional cameras. We consider that the camera receiver is operated at a frame rate (sampling rate) following the Nyquist criterion – at at least 2x the transmission rate. Thus, the LED signals are sampled by the camera such that each transmit bit has at least 2 image frames with at a set of pixels registering a pixel intensity corresponding to that bit's transmit signal intensity. If the sampling is clean, the pixels corresponding to the LED region will register a high pixel intensity when LED transmits a bit 1, and will register a low pixel intensity when the LED transmits a bit 0. Each camera image at each instance of time registers an LED's single state. Hence, for a N bits sequence ID, we consider $2N$ consecutive frames and input to our correlation localization algorithm to identify the LED's exact location on the camera pixel domain.

4.4.1 Correlation Localization Algorithm

We define the localization problem as identifying at least a 3 x 3 pixel block in the camera sampled images that overlaps with the pixels that have registered the LED. The algorithm is setup as a two-phased approach. Phase 1 extracts the data from the images and prepares it for the LED pixel location identification using correlation calculations in Phase 2.

4.4.1.1 Phase 1: Data preparation

Image formatting. Each sampled image at the receiver, regardless of the original resolution, will be resized to VGA resolution (640 x 480 pixels). This is to minimize the image processing computation time. To ensure the transmissions from the LED are not creating disturbing flickering effects, we operate the LEDs at a minimum of 50 Hz which thus requires the cameras to operate with at least 100 frames-per-second (FPS) sampling. Today's off-the-shelf mobile cameras can reach 100 FPS and beyond but at VGA resolution. The images are processed further in gray-scale. We use grayscale version of the sampled color images for post processing only. The camera capture in our experiments is set to capture at full high-definition (1920 x 1080) resolution in RGB color in uncompressed format.

Registering the Images. When the transmitter and/or receiver is in motion, the images sampled at each instance (with 1/FPS seconds separation) may not be aligned spatially. This means that the actual pixel(s) position of the LED will not be the same across successive image frames. To account for this and to ensure the pixel positions of the LED can be spatially overlapped, we register the images using traditional computer vision based image alignment [71] techniques. The alignment is essentially achieved over an image pair, where one image is the reference and the other is the motion frame. The effective 'movement of the scene in pixels'is estimated and corrected (inversed) using a homography (pixel-to-pixel spatial relationship between image pairs) calculations. In our algorithm we take a set of $2N$ (for a N bit ID) consecutive image frames and conducts the image alignment for each sequential pair; that is, (img1, img2) then (img2, img3) and so on. Each pair of aligned images are then virtually superimposed onto the reference image's pixel domain. If the image alignment was ideal, then the LED pixel regions (and other objects in the scene) will precisely overlap. Inefficiencies in practical alignment algorithms can lead to slight mismatches in registration, however, can be considered insignificant as the primarily goal is to overlap as much of the LED pixels across the $2N$ frames with allowance of small errors. An example of the registration using image alignment process for a series of three image pairs is shown in Figure 4.5.

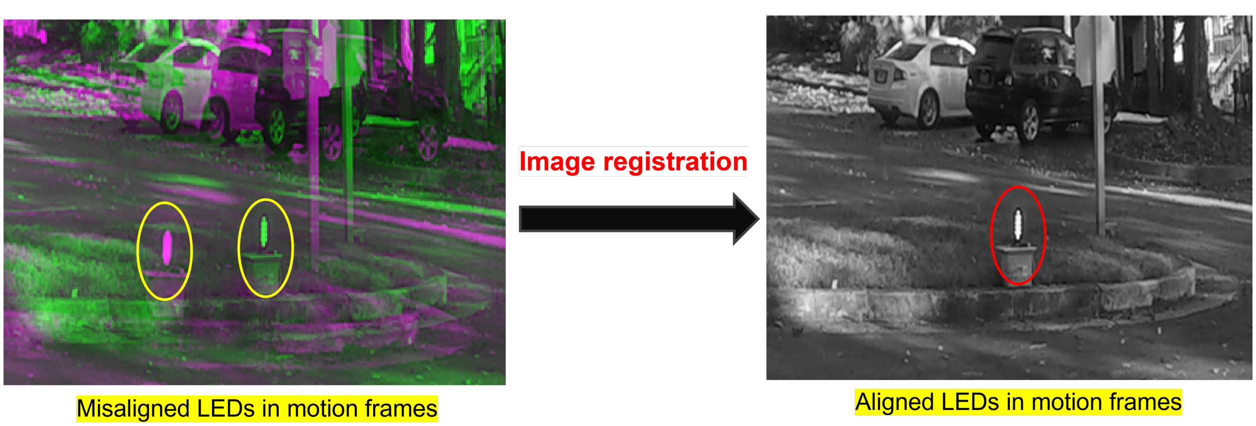


Figure 4.5: Example of misaligned frames in pre-registration (left) and post-registration (right) cases.

4.4.1.2 Phase 2: LED pixel localization using correlation

Correlation. The raw pixel intensity (P) from each pixel coordinate (x, y) from each of the sets of $2N$ images are collected as a single $2N$ element row vector. We prepare another row vector of size $2N$ which contains the N ID sequence bits (I) with every alternate bit as a repetition of an ID sequence bit. These two row vectors are correlated and the effective correlation value is recorded as the *correlation pixel intensity* at the x row and y column of a *correlation image matrix*. We use the definition of cross-correlation between image pixel intensity and bit sequence values as follows:

$$\mathrm{corr}(\vec{P}, \vec{I})[k] = \sum_{m=-\infty}^{\infty} P[m]I[m-k], \tag{4.1}$$

where, $\vec{P}$ represents pixel intensities and $\vec{I}$ represents ID bits and k is the index.

Considering that pixel intensities are non-negative values (for 8 bit gray-scale images the values are in [0 255]), to capture the high-low (ON-OFF) transitions in the bits we consider that the ID correlation vector has 1 mapped to 255 and 0 mapped to -255. This polarity introduction can help filter static pixels from intensity varying pixels. For example, if ID sequence is [1 0 1 1 0], then the correlation ID vector will be [255 255 -255 -255 255 255 255 255 -255 -255]. For the LED region, assuming it were noisy may hold values as [200 195 120 110 210 200 200 198 100 90] compared to a saturated region that will register pixels [255 255 255 255 255 255 255 255 255 255]. The LED pixel to saturated (noise or non-LED) region is 1783:765, which is at least a 2x in differentiation between these regions.

Localization after filtering. Ideally, only the pixels corresponding to the LED in the image will yield higher correlation values compared to other regions. However, in reality, imperfections in image sampling, image artifacts (e.g. blur) and possibility of other things in the scene that look similar to the LED, will result in possibly multiple pixel regions having high correlation values that may be very close to set a general threshold for detection. To address this issue, we first run a correlation and flag the pixels that have high correlation values that are within 10% difference of each other. We set all the other pixels as 0. From this coarse filtered set, we further flag all the pixels which have the least set of variations in their intensity across the $2N$ images. We identify this by setting a 25% gradient threshold for pixel intensity changes across the high to low transitions and vice-versa. We flag the pixels with less than this threshold of variation and set their values as 0, keeping the raw pixel intensity values intact for others. Then we run the correlation calculation for the modified column vectors and choose the pixel(s) with the maximum correlation value (within 1% difference) as the LED pixels.

Unwarping. The registration process essentially warps the set of images to a common pixel domain spatial reference. The LED pixel localization achieved in the previous step should be noted as the LED pixel location on the reference image. The actual LED pixel location on the other images in the set is computed by remapping the pixel coordinates across the registered images using the unwarping process. In this way, through a one-shot correlation process, the LED pixel can be spatial and temporally tracked continuously on each sampled image frame, without any additional computer vision feature extraction.

4.4.2 Assumptions and Potential Solutions

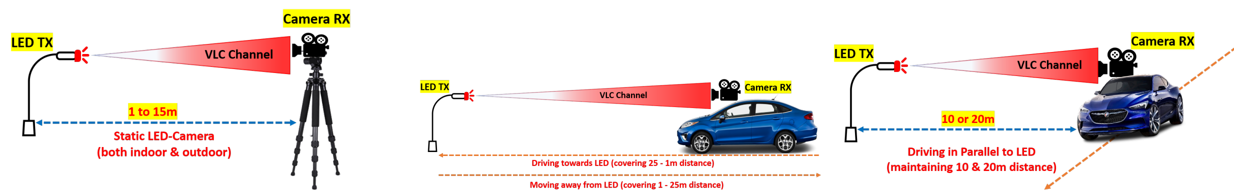
The fundamental assumption in our system is that the camera receiver has knowledge of the dataset of transmission IDs (bit sequences). We justify this assumption using the fact that such knowledge can be generated using multiple techniques depending on the application scenario; (a) the transmitter and receiver can agree apriori on the set of IDs (example use-case: for robot navigation and mapping in finite spaces with small number of LEDs); (b) the LED can transmit, using the VLC channel, a data packet appended to the bit sequence, with the sequence serving for coarse spatial detection of the LED region and the data packet containing the unique ID. The camera receiver can acknowledge reception of the unique ID using a feedback radio channel (example use-case: localizing in a conference setting a large number of mobile devices fit with LEDs); (c) the transmitter and receiver, both, can be connected to a common cloud (wired to infrastructure or cellular) server and commonly be informed on the unique IDs allotted for each LED at a specific location at specific time-slot (example use-case: LEDs attached to buildings or road infrastructure and camera on vehicles used for scene perception).

4.5 Implementation and Evaluation

We evaluate the performance of the optical correlation based localization method through a experimental trace-based analysis. We setup a LED and camera in indoor (home) and outdoor settings, and conducted experiments by varying different parameters in each experimentation trial, and collected data traces. Each data trace or sample is a camera image

(a) (b) (c) (d)

Figure 4.6: Experimental setup samples for LED-Camera communication link (a) indoor (enclosed room environment), (b) outdoor (open space and parking lot in an apartment complex area) (c) driving towards LED, and (d) driving parallel to LED.



(a) LED-Camera communication link in static setups for both indoor and outdoor

(b) LED-Camera communication link in car forward and reverse driving scenarios

(c) LED-Camera communication link in car driving parallel to LED scenarios (10 and 20 m distance)

Figure 4.7: Illustration of LED and camera setup in our experiments for, (a) static indoor and outdoor, (b) driving towards or moving away from LED, and (c) driving parallel to LED (passing the LED on right side of the car driving direction) cases

frame of a video footage recorded at specific resolution and video capture frame-rate. In our evaluations, we consider a single LED and a single camera setup, where we used a solid-state 1 Watt LED for indoors and a 10 Watt brake/trail light LED for outdoors, both modulated at 60 Hz. We used a GoPro Hero 6 as the camera set at 120 frames-per-second. Each trace of our experiments was 1min long footage. Overall, our dataset for LED localization evaluation contains about 15000 non-repetitive (LED location on each frame differs from other by at 3-5 pixels) images. We evaluate our system across four different real-world LED-camera settings (Figure 4.6, under static and motion configurations (Figure 4.7). As a default, we use ‘10110

'($N = 5$) as the bit (ID) sequence and $2N = 10$ frames for correlation. Processing was done on data traces, with analysis conducted offline using MATLAB on an Intel i7 laptop PC.

1. **Static-Indoor**: indoor room environment enclosed by walls with a LED and camera set at specific distance in static position. The distances evaluated include 1 m, 2 m and 3 m. The room was well lit with artificial ceiling lighting and in some cases sunlight from windows and doors.

2. **Handheld Camera Motion-Indoor**: same setup as in static-indoor but with camera being hand-held and panned from left to right of the LED in view. Distances evaluated include 1 m, 2 m and 3 m.

3. **Static-Outdoor**: In an open outdoor space of apartment complex where the sunlight is abundant, LED was placed on ground with some slight elevation using a mount and camera on a tripod. Distances evaluated include 5 m, 10 m and 15 m.

4. **Driving Motion-Outdoor**: In the outside parking area of apartment complex, same setup as the static-outdoor but with camera mounted on a tripod, hand-held, with the experimenter in the passenger seat of a driven car. Keeping the LED within the field-of-view (FOV) of the camera, the car was driven along the following trajectories: (a) Drive towards LED at 10-20 mph car speed and distance variable from 25 m to 1 m, (b) Drive away from LED at 10-20 mph car speed and distance variable from 1-25 m, (c) Drive parallel to the LED (passed the LED on the right side of the car's driving

direction) at a distance of 10 m, and (d) Drive parallel to the LED at a distance of 20 m.

We evaluate the performance of our localization method using the *average localization accuracy* as the metric; defined as the ratio of the *total number of camera image frames with successful localization* to the *total number of image frames in the data trace*, averaged across multiple experimentation trials. We define a *successful localization* as when the localization algorithm detects at least one (non-overlapping) 3 x 3 pixel region that intersects with the LED region-of-interest (ROI). An LED ROI is the rectangular pixel region that completely houses the LED in the particular camera image.

This heuristic choice of 3 x 3 pixel ROI corresponds to a *strict* threshold for the localization accuracy evaluation. It is common practice in computer vision analysis to require any detection ROI be larger than a 1 x 1 pixel. This creates a trade off – large ROI leads to more outliers and strict ROI can lead to low detection accuracy. However, we chose to use a strict threshold of 3 x 3 pixels in our evaluation, at a processing resolution of 640 x 480 pixels. We recall our mention from the earlier section that, regardless of the camera capture resolution, we convert all image frames to 640 x 480, to standardize the processing method as well as optimize for real-time performance.

The ROI will change with the distance between the camera and LED; at shorter distances the ROI will be larger thus providing a larger number of ROI intersecting 3 x 3 pixel regions, which significantly reduces as the distance increases. For example, the ROI of the brake/trail light LED at 5, 10, 15 and 20 m on the GoPro camera at VGA resolution are listed in

Distance [m]	ROI [pixels]
5	24 x 50
10	18 x 31
15	14 x 27
20	10 x 15

Table 4.1: Camera pixel ROI of LED at different distances experimented in our outdoor evaluation. The ROI is the set of pixels over a rectangular region in the image where all the pixels in the ROI encompass the LED. The pixels that correspond to the partial registration of the LED due to the curvature of the LED shape are not considered in the ROI.

Table 4.1. We observed that even at 20m range, there are at least three 3 x 3 non-overlapping LED regions that can be marked for localization of the LED.

In summary, we include the following evaluation results,

1. Comparative evaluation of localization accuracy of our optical correlation method in indoor and outdoor and under static and motion cases. We compare with (i) LED detection using color based thresholding, (ii) computer vision based technique that uses aggregate channel features (ACF) and support vector machine (SVM) machine learning, and (iii) a customized version of convolutional neural network (CNN) based YOLO v3 deep learning object recognition model.

2. Micro-benchmark evaluation of our optical correlation method across variable, (i) distance between LED and camera, (ii) number of images used for correlation, and (iii) car speed variation in localization accuracy.

4.5.1 Comparative evaluation

We compare the localization accuracy of our optical correlation localization method with traditional techniques. In particular, we consider color thresholding as a basic technique

Experimental Setup	Color Thresholding(%)	ACF based ML Detector(%)	YOLO v3 indoor custom trained on indoor data(%)	YOLO v3 outdoor custom trained on outdoor data(%)	YOLO v3 combined custom trained on complete dataset(%)	our Optical Correlation algorithm(%)
Static, Indoor [1m, 2m, 3m]	69.85	76.0	49.25	0.09	63.78	100
Motion, Indoor (Hand held) [1m, 2m, 3m]	63.68	73.33	81.91	0.60	92.65	94.76
Static, Outdoor [5m, 10m, 15m]	68.73	67.5	3.57	97.34	92.78	98.08
Motion, Outdoor (Driving) [Forward, Reverse, Parallel 10m and 20m]	37.78	62.5	0.10	85.84	68.92	86.69

Table 4.2: Average localization accuracy metric based comparative evaluation of optical correlation localization with color thresholding, ACF-ML detector and YOLO v3-Deep learning classifier.

Experimental Setup	ACF ML			YOLO v3 indoor			YOLO v3 outdoor			YOLO v3 combined			Optical correlation		
	avg. P	avg. R	avg. F1	avg. P	avg. R	avg. F1	avg. P	avg. R	avg. F1	avg. P	avg. R	avg. F1	avg. P	avg. R	avg. F1
Static, Indoor	0.76	0.5	0.59	0.73	0.69	0.71	0.05	0.01	0.01	0.81	0.77	0.79	1.0	1.0	1.0
Motion (hand held), Indoor	0.73	0.50	0.61	0.84	0.86	0.85	0.23	0.01	0.02	0.95	0.95	0.95	0.89	1.0	0.94
Static, Outdoor	0.67	0.73	0.69	0.29	0.07	0.12	0.98	0.98	0.98	0.95	0.95	0.95	0.98	1.0	0.98
Motion (Driving), Outdoor	0.62	0.72	0.68	0	0	Nan	0.89	0.90	0.90	0.81	0.83	0.82	0.86	1.0	0.92

Table 4.3: Localization average precision (P), recall (R) and F1-score metric based comparative evaluation of optical correlation localization with color thresholding, ACF-ML detector and YOLO v3-Deep learning classifier. True Positive (TP) is when an LED location is accurately localized for a given frame. False Positive (FP) is when the LED is not present in the scene and but the system provides an erroneous LED localization output. True Negative (TN) is when the system reveals there is no LED when there is no LED actually. False Negative (FN) is when the system reveals LED localized pixels when there is no LED actually. To serve as Negative data, we captured images in different experiment settings used for our evaluation, without the LED transmitter.

typically used in detection processes using computer vision. Next, we consider a more advanced feature based LED detection technique called ACF detector that marks a set of structural features on the object. The features are then set to learn using a SVM machine learning model. Finally, we compare with state-of-the-art deep learning classification techniques, particularly, with YOLO v3 that essentially functions as a single-shot classifier.

4.5.1.1 Baseline for comparison

In each of the comparative methods we use the traditional implementations and make slight modifications to fit out experimentation to set a common baseline for evaluation.

Color thresholding. Considering the color of the LED is more in the RED space, we set a threshold for the average intensity of the pixel to be detected as an LED. We calibrate the threshold for each experiment trace by selecting the average intensity of the HIGH (LED ON) and LOW (LED OFF) pixels across the images in each 1 min trace.

Machine learning with aggregate channel features (ACF) [112]. This method is a supervised machine learning approach. ACF detector uses an effective sliding window detector to extract the variations in the structural features in the scene. During data labeling, we labeled by specifying a bounding box region for the LED region in each image. The outcomes of the ACF detector is the estimated LED detection region of pixels. The intersection over union (IoU) for the region is set to 0.5 (50%).

Deep learning with YOLO v3 [113]. YOLO v3 (you only look once, version 3) is a state-of-the-art CNN model which uses 1x1 convolution layers for prediction, and is traditionally trained on MSCOCO dataset which contains 80 object categories. However, the MSCOCO dataset does not contain LED images in ON/OFF state. In this regard, we created a LED dataset by labeling over 15000 images of both LED ON and LED OFF states, equally distributed. The dataset houses LED images from 1. Static Indoor scenario, 2. Motion Indoor scenario, 3. Static Outdoor scenario and 4. Motion outdoor scenario to accommodate all the variations and serve as a representative of the real world scenarios. The labeling of the dataset was carried out using an open source tool, labelImg [114] and labels were exported in the desired YOLO format. To train the model on the custom dataset, a transfer learning approach was adopted. YOLO v3 uses a variant of Darknet, which originally has 53 layer network trained on Imagenet [115]. For the task of detection, 53 more layers are stacked onto it along with residual skip connections, and upsampling layers, forming a 106 layer fully convolutional underlying architecture for YOLO v3. The pre convolutional weights of darknet53-conv74 were used to train the custom YOLO model where the weights of darknet 53 model with pre convolutional weights were used for the initial 74 layers the rest were are trained from scratch on the data set we have collected. We considered three types of evaluation for the YOLO v3 model used for evaluation. First, we trained entirely on the indoor images and tested on the same. Next, we trained on the outdoor images and tested on the same. Last, we trained on the entire dataset and tested on the entire dataset. We used 60:40 distribution for training:test sets, and randomized the test-set for total 5 trials. We computed the average of the localization accuracy across such an evaluation.

4.5.1.2 Results

We summarize the performance of our approach compared with the baseline techniques using average localization accuracy metric in Table 4.2. We observe that our optical correlation technique outperforms the comparative techniques in general. We make the following specific observations from the evaluation results:

- We observe that the localization accuracy of our approach is relatively lesser in motion cases. Upon analysis we learned that the localization errors in motion cases are primarily due to the errors in the image registration process, which may not necessarily be 100% accurate. However, even with the a simple off-the-shelf image registration technique used in computer vision, our algorithm outperforms the comparative techniques.

- The comparative techniques perform poorly in locating the LED from the scenes, especially for those frames where the LED is in 'OFF' state. Extracting LED locations in 'OFF' frames is challenging, as LED in general is not a feature-rich object. The LED OFF state further adds to the challenge as the intensity of the pixel region is very low and thus making geometric and photometric feature dependent analysis, such as color thresholding and ACF, very challenging. The lack of features fails to effectively train the YOLO v3 deep learning model for LED OFF states.

- The YOLO v3 deep learning model performs the best when it is trained and tested across the entire dataset. When trained and tested on a specific setting such as only indoor or only outdoor, the model performs poorly. This is attributed to the lack of variations in features across the dataset which limits the learning process efficiency. We observe that there is no clear insight that can be gained about the learning process of YOLO v3 for LED detection as the accuracy numbers do not necessarily follow any trend. In this work, we setup a baseline deep learning LED recognition, which shows some potential, however, not better than optical correlation. We posit that these evaluation results reveal the need to further explore machine/deep learning models for LED localization. We have provided some examples of success and failure cases of

YOLO v3 LED localization performance in Appendix A.

- We also present the average precision, average recall, and average F1-score values for our evaluation in Table 4.3. The fidelity of the optical correlation method is reflected in its high average recall values and F1- scores.

4.5.2 Microbenchmarks

4.5.2.1 Distance between transmitter and receiver

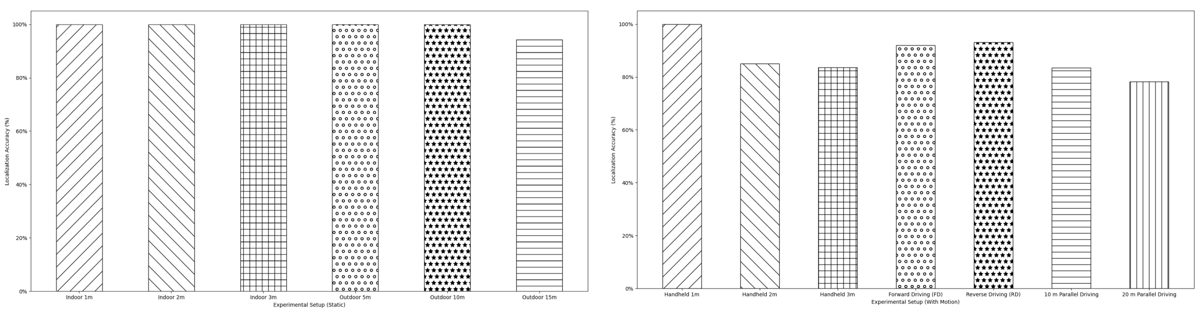


Figure 4.8: Average localization accuracy of optical correlation in indoor and outdoor, (i) static (left), and (ii) motion (right).

Figure 4.9: Outdoor static experimental setup with trail light LED (left) placed in a shaded area (right) placed in a bright spot where sunlight directly falls onto the LED.

From Figure 4.8, it can be observed that for both indoor and outdoor static cases, the average LED locating accuracy is about perfect, when the LED-Camera distance is up to

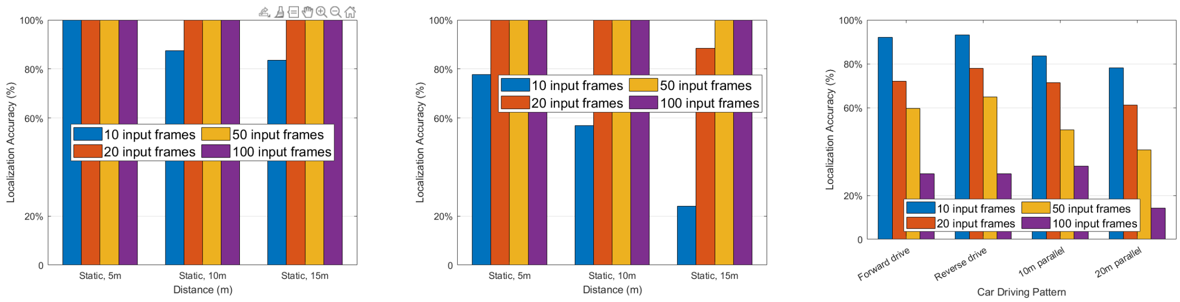


Figure 4.10: Number of Input frames in Correlation Vs Accuracy Analysis for (left) Outdoor Static & LED is at shaded spot, (middle) Outdoor Static & LED is at bright spot, (right) Outdoor Motion (Car Driving) cases.

10 m. However, at 15 m distance in outdoor experiment, the average accuracy is about 94%. In outdoor setup, when the LED is kept in a spot where the sun/ambient light shines bright on the LED (right image of Figure 4.9), due to the presence of saturated regions in the image which do not correspond to the LED. The intensity changes in the ON and OFF patterns will be impacted leading to detection errors. Such saturated regions might have higher correlation values under optical correlation leading to LED localization outliers. We consider both bright and shaded spot outdoor setup (shown in Figure 4.9) with the variations of distance in our analysis, and present the accuracy results in Figure 4.8. We report that with 20 input frames, when LED is placed at bright spot at 15 m distance, the average accuracy is about 88.5% and with the same specifications, at shaded spot the LED can be almost perfectly localized. These results clearly explain the impacts of LED-Camera distance and sunlight reflections on optical correlation localization accuracy.

4.5.2.2 Number of input frames for correlation

For all the outdoor static setups, we also test our system by varying the number of input image frames (from 10 to 100 images) in each execution of the correlation and report the results in Figure 4.10. We consider both, shaded and bright spot LED, cases while changing the correlation input frames in our analysis and show both results in left and middle illustrations of the Figure 4.10. We notice that with the increase in the number of input frames during correlation, the optical correlation method reaches near-perfection, even when LED is kept in extreme bright spot scenarios. Having more images during correlation helps generate a robust correlation value that can be easily delineated from outliers as there are more bits (values) being multiplied in the cross correlation process. Also, with larger number of images the chances that the scene can precisely mimic the variations in the ON/OFF (1/0) intensities become lower. In particular, we observe that with 10 frames, the accuracy is sub-par especially at distances beyond 10 m and static cases. However, just by increasing the input frames to 20, the accuracy can be significantly improved. In contrary to the characteristics and results of the static experiments, in motion driving cases, accuracy is higher when the number of input image frames in correlation is smaller. We report this behavior for all four driving patterns in Figure 4.10 (right). Under motion, the smaller the number of frames being considered for alignment is better as the amount of actual physical motion in the scene may be (almost insignificant) low. For example, 10 frames at 120 FPS is about 9 ms time span. The amount of motion that can happen within such a duration is typically low, except when the vehicle is driven at highway speeds. We observed from our analysis that the drop in accuracy with increasing frames is primarily due to registration errors, which is in turn a function of vehicle speed.

4.5.2.3 Car driving speed variation in localization accuracy

To evaluate our system performance in LED localization for outdoor motion cases, we extend the experimentation with the variation of car driving speed from 5 mph to 30 mph towards the LED emitter and include the results in Figure 4.12. By placing the LED transmitter as static on a tripod stand, we drive the car towards the LED attaching the camera on the wing (side-view) mirror of the car, as shown in Figure 4.11. We observe that the average LED localization accuracy is about 98% while driving the car at 5 mph and is about 87% when the car speed increases to 30 mph. As we mentioned earlier, the localization accuracy might be lesser in motion cases due to the dependency on the image registration performance. With higher driving speed, the movements in pixels are also greater compared to the static or slow driving cases. So, the misalignment still exists even after registering the motion frames. As shown in Figure 4.13, the misalignment in registered frames also increases when the car drives faster (30 mph) compared to a slower speed (5 mph). We observer that such misalignments are fairly small and are within the range that can be handled by state-of-the-art camera motion stabilization, such as by inverting the motion artifacts using motion vectors generated by inertial measurement units (IMU) or using computer vision optical flow methods. We target to incorporate such techniques in our future work.

Figure 4.11: Outdoor setup: Driving towards LED transmitter (kept at static) with varying the speed of the car from 5 mph to 30 mph.

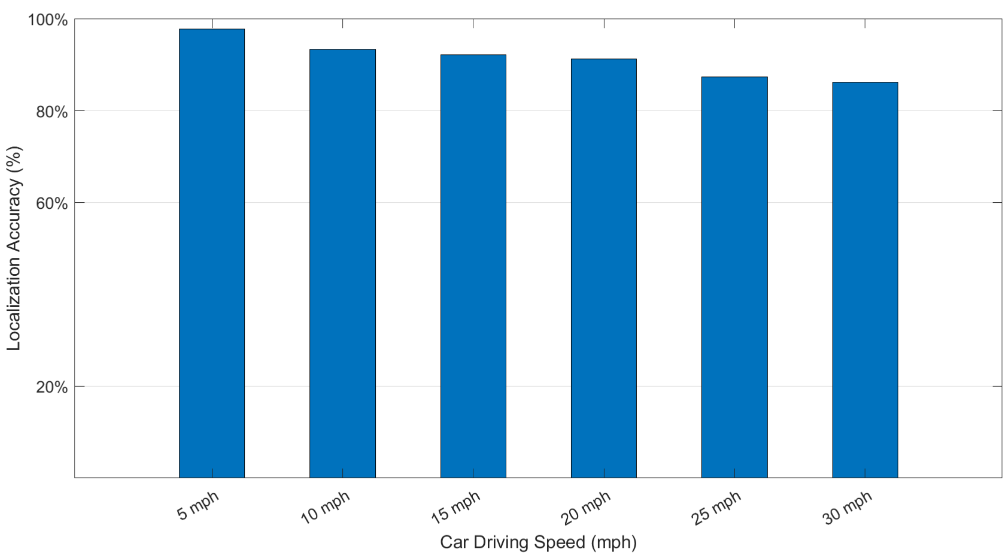
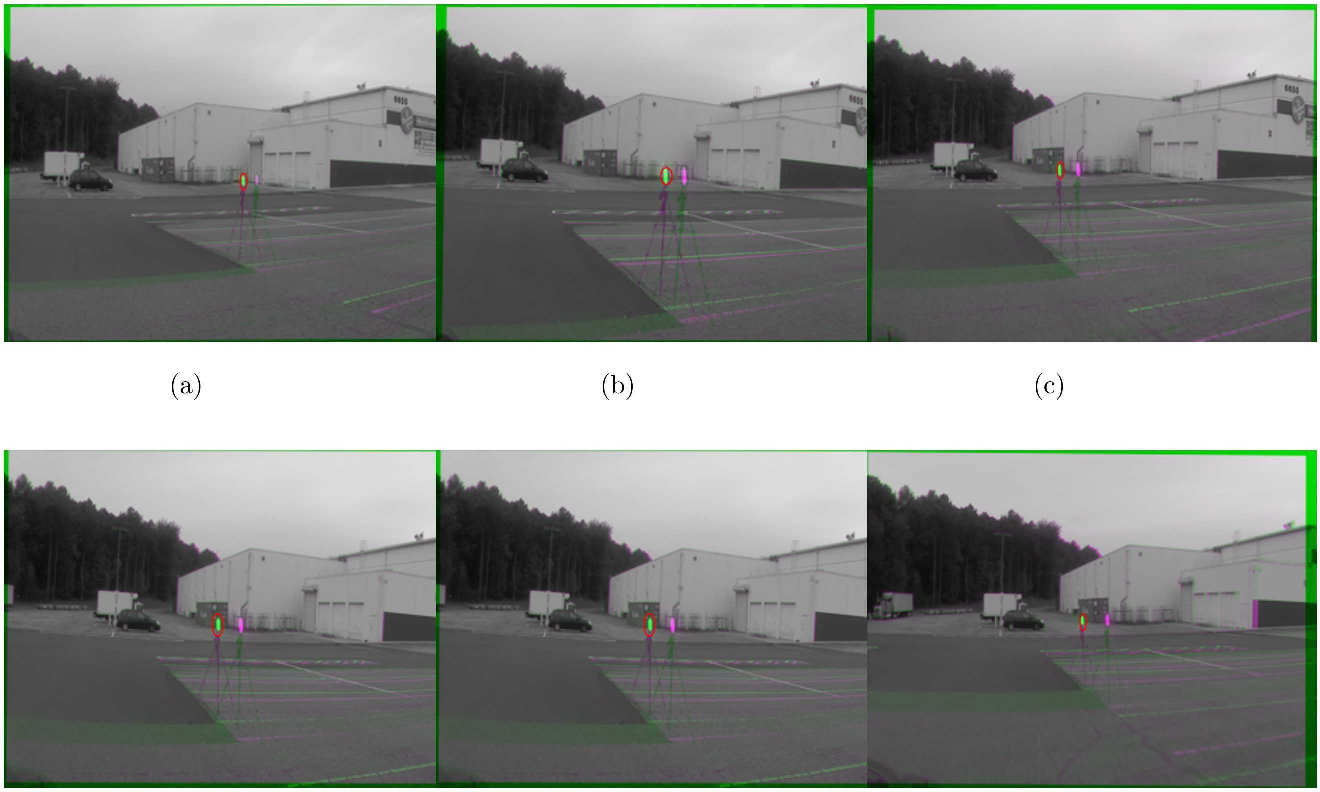


Figure 4.12: LED localization accuracy with the variation of car driving speed.

4.5.2.4 Timing analysis of correlation algorithm

We present the execution time of each of the steps in our algorithm in Table 4.4. In static cases, the algorithm does not require to implement image registration and hence it performs



(a) (b) (c)

(d) (e) (f)

Figure 4.13: Illustration to show how the misalignment due to the movement of car at different driving speeds (a) 5 mph, (b) 10 mph, (c) 15 mph, (d) 20 mph, (e) 25 mph, (f) 30 mph, reduces the LED localization accuracy of our correlation algorithm, even after implementing proper image registrations (state-of the-art).

faster than the motion cases. We report that our algorithm takes on average 0.29 seconds to locate the LED in each of the inputs of motion frames and can process each of the static images within the average of 0.22 seconds. We also compared the average LED localization processing time for each images of our algorithm with the other techniques which are used in our baseline localization performance comparison. We report each execution time of the implemented algorithms in Table 4.5 and notice that our correlation algorithm takes less time to process compared to simple color thresholding and machine learning based techniques. However, we do note that our technique, though slower than YOLO v3 based LED detection, we recall that YOLO v3 has much lower localization accuracy. This creates a trade-off between computation versus accuracy, and we hypothesize that future works could use a hybrid method that integrates YOLO v3 with correlation to achieve the best of both worlds.

Correlation algorithm steps	**Average time taken (seconds)**
Reading image inputs (10 frames)	0.337477
Image registration (motion cases)	0.725
LED ID extraction	0.000906
Image correlation with ID	0.827190
Set threshold and decision making	0.874364
Image un-warping and saving coordinates	0.140766
Total processing time (10 input images)	2.90
LED localization time for each input image (motion)	0.29
LED localization time for each input image (static)	0.22

Table 4.4: Timing analysis for each steps in our correlation algorithm to locate LED on the input image frames.

Implemented System	**Average time taken (seconds)**
Color Thresholding	1.043
ML with ACF	0.538
DL with YOLO v3	0.018
Correlation algorithm	0.29

Table 4.5: Comparing average execution time (seconds) to locate the LED emitter in each of the input image frames of different algorithms.

4.6 Conclusion

We designed a novel optical correlation based localization to precisely and accurately locate LED emitters in camera images. We designed and implemented the optical correlation algorithm and evaluated using real-world experiment traces. Upon evaluation in indoor, outdoor, static and motion cases, and comparing with traditional ML and non-ML techniques for LED detection, we showed optical correlation outperforms the comparative techniques.

We showed that traditional feature based techniques fail due to lack of features in LED image regions. We learned from the evaluation that our optical correlation technique's localization accuracy has a trade off in static and driving cases for the choice of the number of input correlation frames. Our evaluation also revealed state-of-the-art classification using YOLO v3 deep learning does not necessarily solve the problem as the training process does not reveal any evidence that the model is able to learn unique characteristics about the LEDs. We posit that further exploration in optical correlation assisted deep learning models may be useful for improving optical camera reception fidelity, particularly in visible light and camera communication applications.

We note that scalability is a problem when it comes to creating unique blinking sequences for each LED in the field-of-view of the camera. We note to the reader that this can be resolved by using a finite set of sequences and reusing the sequences, but at different frequencies and different dynamic ranges (difference between ON and OFF intensities). The scalability question generates an interesting problem of *recognizing* the LED after it has been detected. We propose that we can use uniqueness in ID (sequence), frequency and intensity as parameters, which can overall, scale the number of options considering the number of permutations possible. Further, it is possible to use contextual relevance of the LEDs – what are they attached to and what are the objects/entities detected and recognized in vicinity – rely on state-of-the-art computer vision object detection. We believe our current results present a foundation for the future work that can incorporate such and variations of techniques for addressing scalability.

CHAPTER 5
CONCLUSIONS

With the growing interest in VLC systems research and development of optical wireless technologies and standards, it is evident that VLC and the notion of next generation wireless technology go hand-in-hand. However, realizing VLC as a front-end commercialized product requires significant advancements which need to primarily cater addressing its fundamental challenges, such as noise, interference, and LOS degradation during mobility. This thesis has focused on addressing these challenges.

The fundamental notion of this thesis is to improve the overall signal reception quality in VLC systems using spatial dimensions of light signals. In doing so, through this thesis, my work has designed, prototyped, and evaluated a novel pixelated shutter-based VLC receiver based on spatial filtering. This filtering mechanism can isolate signal from noise and interference signals which improves signal to noise Ratio (SNR) ensuring higher data rates in VLC. Then, we developed a fast spatial tracking mechanism to identify the location of the signal on the receiver under mobility. This design included a single photodiode based VLC receiver and successfully demonstrated multiple access. This way, the system performance was enhanced by allowing more signals from different spatial locations at the same time. Finally, we introduced a camera based visual identification solution to detect and track LED transmitter for VLC enabled applications. Such LED localization can significantly assist in autonomous driving or vehicular VLC based applications.

In conclusion, throughout my doctoral research journey, I have striven to address several fundamental challenges in VLC and VLC based applications and the core contributions or outcomes of my research efforts have been published and presented in top tier conferences and journals. I believe, my dissertation will set a foundation step towards designing next generation high speed, robust, reliable, and mobile visible light communication systems.

REFERENCES

[1] Vlc band image. `https://www.once.lighting/visible-light-spectrum/`.

[2] T. Komine and M. Nakagawa. Fundamental analysis for visible-light communication system using led lights. *IEEE Transactions on Consumer Electronics,*, 50(1):100–107, Feb 2004.

[3] VLC. Visible light communication consortium. Available online, April 2018.

[4] VLC. Augmented reality using vlc. Available online, August 2017.

[5] Phillips. Philips indoor vlc positioning technology. Available online, August 2018.

[6] Under Water VLC. Firefly underwater vlc. Available online, August 2018.

[7] Saeed Ur Rehman, Shakir Ullah, Peter Han Joo Chong, Sira Yongchareon, and Dan Komosny. Visible light communication: a system perspective—overview and challenges. *Sensors*, 19(5):1153.

[8] Murat Uysal, Carlo Capsoni, Zabih Ghassemlooy, Anthony Boucouvalas, and Eszter Udvary. *Optical wireless communications: an emerging technology.* Springer, 2016.

[9] A. Ashok. Doctoral dissertation: Design, modeling and analysis of visual mimo communication. In *RUcore: Rutgers University Community Repository*, 2014. NJ, USA.

[10] S. Perli, N. Ahmed, and D. Katabi. Pixnet: Lcd-camera pairs as communication links. In *SIGCOMM*, pages 451–452, 2010.

[11] A. Ashok, S. Jain, M. Gruteser, N. Mandayam, Wenjia Yuan, and K. Dana. Capacity of pervasive camera based communication under perspective distortions. In *Pervasive*

Computing and Communications (PerCom), 2014 IEEE International Conference on, pages 112–120, March 2014.

[12] Ankit Mohan, Grace Woo, Shinsaku Hiura, Quinn Smithwick, and Ramesh Raskar. Bokode: imperceptible visual tags for camera based interaction from a distance. In *SIGGRAPH '09: ACM SIGGRAPH 2009 papers*, pages 1–8, New York, NY, USA, 2009. ACM.

[13] Barry D Van Veen and Kevin M Buckley. Beamforming: A versatile approach to spatial filtering. *IEEE assp magazine*, 5(2):4–24, 1988.

[14] M Stojanovic, JA Catipovic, and JG Proakis. Spatial signal processing for time-dispersive digital communication channels. *IFAC Proceedings Volumes*, 27(8):1257–1262, 1994.

[15] M Stojanovic, JA Catipovic, and JG Proakis. Reduced-complexity spatial and temporal processing of underwater acoustic communication signals. *The Journal of the Acoustical Society of America*, 98(2):961–972, 1995.

[16] M. R. Rahman, S. M. T. Islam, and A. Ashok. A novel architecture for ultra-high signal-to-interference-noise-ratio reception in visible light communication. In *2019 11th International Conference on Communication Systems Networks (COMSNETS)*, pages 172–179, Jan 2019.

[17] Rashed Rahman and Ashwin Ashok. High throughput mobile visible light communication. In *2020 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops)*, pages 1–2. IEEE, 2020.

[18] MD Rahman, Kehinde Adedara, and Ashwin Ashok. Enabling multiple access in visible light communication using liquid crystal displays: A proof-of-concept study. *Electronics*, 9(5):826, 2020.

[19] Md Rashed Rahman, TV Sethuraman, Marco Gruteser, Kristin J Dana, Shubham Jain, Narayan B Mandayam, and Ashwin Ashok. Camera-based light emitter localization using correlation of optical pilot sequences. *IEEE Access*, 10:24368–24382, 2022.

[20] NASA. Light technology being developed for advanced communications. Available online, August 2017.

[21] T. Komine and M. Nakagawa. Integrated system of white led visible-light communication and power-line communication. *IEEE Transactions on Consumer Electronics,*, 49(1):71–79, Feb. 2003.

[22] IEEE. Ieee 802.15.7 wpan task group 7: Visible light communication. Available online, April 2018.

[23] LiFi. purelifi: The lifi technology. Available online, September 2021.

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

[25] A. Jovicic, J. Li, and T. Richardson. Visible light communication: opportunities, challenges and the path to market. *IEEE Communications Magazine*, 51(12):26–32, December 2013.

[26] IEEE. Ieee 802.15.7r1 status, 2015. Available online, March 2018.

[27] Andrea Goldsmith. *Wireless Communications*. Cambridge University Press, 2005.

[28] S. Rajbhandari, A. V. N. Jalajakumari, H. Chun, G. Faulkner, K. Cameron, D. Tsonev, H. Haas, E. Xie, J. J. D. McKendry, J. J. D. Herrnsdorf, R. Ferreira, E. Gu, M. D. Dawson, and D. C. O'Brien. A multi-gigabit/sec integrated multiple input multiple output vlc demonstrator. *Journal of Lightwave Technology*, PP(99):1–1, 2017.

[29] Sony. Sony's 1000fps camera. Available online, March 2018.

[30] A. T. Hussein and J. M. H. Elmirghani. 10 gbps mobile visible light communication system employing angle diversity, imaging receivers, and relay nodes. *IEEE/OSA Journal of Optical Communications and Networking*, 7(8):718–735, August 2015.

[31] Boston University. Smart lighting research center. Available online, March 2017.

[32] N. O. Tippenhauer, D. Giustiniano, and S. Mangold. Toys communicating with leds: Enabling toy cars interaction. In *2012 IEEE Consumer Communications and Networking Conference (CCNC)*, pages 48–49, Jan 2012.

[33] H. Ma, L. Lampe, and S. Hranilovic. Integration of indoor visible light and power line communication systems. In *2013 IEEE 17th International Symposium on Power Line Communications and Its Applications*, pages 291–296, March 2013.

[34] Jiangtao Li, Angli Liu, Guobin Shen, Liqun Li, Chao Sun, and Feng Zhao. Retro-vlc: Enabling battery-free duplex visible light communication for mobile and iot applications. In *Proceedings of the 16th International Workshop on Mobile Computing Systems and Applications*, HotMobile '15, pages 21–26, New York, NY, USA, 2015. ACM.

[35] Junrui Yang Chenren Xu Guobin Shen Guojun Chen Xieyang Xu, Yang Shen and

Yunzhe Ni. Passivevlc: Enabling practical visible light backscatter communication for battery-free iot applications. In *Proc. ACM Mobile Computing and Networking (MobiCom)*, Oct 2017.

[36] Liquid Crystal Technologies. Lcd shutters from liquid crystal technologies. Available online, March 2018.

[37] Ettus. Ettus research: Usrp. Available online, March 2018.

[38] DigiKey. Sfh 206 k photodiodes. Available online, March 2018.

[39] DigiKey. Red led ltl2r3ew3ks. Available online, March 2017.

[40] KNORVAY. Wireless laser pointer. Available online, March 2018.

[41] 4D Systems. Lcd shutters from 4d systems. Available online, March 2017.

[42] Keithley. 2231a-30-3 triple channel dc power supply. Available online, March 2018.

[43] Tektronix. Mdo4000c mixed domain oscilloscope. Available online, March 2018.

[44] RaspberryPi.org. Camera module v2 from raspberrypi.org. Available online, March 2018.

[45] Texas Instruments. Lm358 n dual operational amplifier. Available online, March 2018.

[46] Edmund Optics. Aspheric condenser lens. Available online, March 2018.

[47] J.M. Kahn and J.R. Barry. Wireless infrared communications. *Proceedings of the IEEE*, 85(2):265–298, Feb 1997.

[48] A. Galisteo, D. Juara, and D. Giustiniano. Research in visible light communication systems with openvlc1.3. In *2019 IEEE 5th World Forum on Internet of Things (WF-IoT)*, pages 539–544, April 2019.

[49] M. Haus, A. Y. Ding, and J. Ott. Localvlc: Augmenting smart iot services with practical visible light communication. In *2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks" (WoWMoM)*, pages 1–9, June 2019.

[50] Shengrong Yin, Nour Smaoui, Milad Heydariaan, and Omprakash Gnawali. Purple vlc: Accelerating visible light communication in room-area through pru offloading. In *EWSN*, pages 67–78, 2018.

[51] C. Liu, X. Jin, W. Zhu, M. Jin, and Z. Xu. Demonstration of a low complexity arm-based indoor vlc transceiver under strong interference. In *2017 13th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pages 622–627, June 2017.

[52] S. Yin and O. Gnawali. Towards embedded visible light communication robust to dynamic ambient light. In *2016 IEEE Global Communications Conference (GLOBECOM)*, pages 1–6, Dec 2016.

[53] H. Marshoud, V. M. Kapinas, G. K. Karagiannidis, and S. Muhaidat. Non-orthogonal multiple access for visible light communications. *IEEE Photonics Technology Letters*, 28(1):51–54, Jan 2016.

[54] C. Chen, W. Zhong, H. Yang, and P. Du. On the performance of mimo-noma-based visible light communication systems. *IEEE Photonics Technology Letters*, 30(4):307–310, Feb 2018.

[55] M. W. Eltokhey, M. Khalighi, and Z. Ghassemlooy. Multiple access techniques for

vlc in large space indoor scenarios: A comparative study. In *2019 15th International Conference on Telecommunications (ConTEL)*, pages 1–6, July 2019.

[56] Ioannis Pefkianakis, Yun Hu, Starsky H.Y. Wong, Hao Yang, and Songwu Lu. Mimo rate adaptation in 802.11n wireless networks. In *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, MobiCom '10, pages 257–268, New York, NY, USA, 2010. ACM.

[57] S. M. Khan, N. Saha, M. R. Rahman, M. Hasan, and M. S. Rahman. Performance improvement of mimo vlc using v-blast technique. In *2016 5th International Conference on Informatics, Electronics and Vision (ICIEV)*, pages 45–49, May 2016.

[58] A. H. Azhar, T. Tran, and D. O'Brien. A gigabit/s indoor wireless transmission using mimo-ofdm visible-light communications. *IEEE Photonics Technology Letters*, 25(2):171–174, Jan 2013.

[59] Y. Zhu, W. Liang, J. Zhang, and Y. Zhang. Space-collaborative constellation designs for mimo indoor visible light communications. *IEEE Photonics Technology Letters*, 27(15):1667–1670, Aug 2015.

[60] M. Ijaz, D. Tsonev, J. J. D. McKendry, E. Xie, S. Rajbhandari, H. Chun, G. Faulkner, E. Gu, M. D. Dawson, D. O'Brien, and H. Haas. Experimental proof-of-concept of optical spatial modulation ofdm using micro leds. In *2015 IEEE International Conference on Communication Workshop (ICCW)*, pages 1338–1343, 2015.

[61] J. Lian and M. Brandt-Pearce. Multiuser mimo indoor visible light communication system using spatial multiplexing. *Journal of Lightwave Technology*, 35(23):5024–5033,

Dec 2017.

[62] K. Kobayashi, H. Kobayashi, and K. Iwashita. Multi-user mimo in visible light communications by superposed intensity modulated signals. In *2019 24th OptoElectronics and Communications Conference (OECC) and 2019 International Conference on Photonics in Switching and Computing (PSC)*, pages 1–3, July 2019.

[63] Y. Xiao, Y. Zhu, D. Zhang, and H. Zhang. High-speed visible light communication chipset based multi-color mimo system. In *2019 28th Wireless and Optical Communications Conference (WOCC)*, pages 1–5, May 2019.

[64] T. V. Pham, H. L. Minh, and A. T. Pham. Multi-cell vlc: Multi-user downlink capacity with coordinated precoding. In *2017 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 469–474, May 2017.

[65] L. Grobe, A. Paraskevopoulos, J. Hilt, D. Schulz, F. Lassak, F. Hartlieb, C. Kottke, V. Jungnickel, and K. Langer. High-speed visible light communication systems. *IEEE Communications Magazine*, 51(12):60–66, December 2013.

[66] S. Chen and C. Chow. Color-shift keying and code-division multiple-access transmission for rgb-led visible light communications using mobile phone camera. *IEEE Photonics Journal*, 6(6):1–6, Dec 2014.

[67] Hyunchae Chun, Ariel Gomez, Crisanto Quintana, Weida Zhang, Grahame Faulkner, and Dominic O'Brien. A wide-area coverage 35 gb/s visible light communications link for indoor wireless applications. *Scientific reports*, 9(1):1–8, 2019.

[68] Emil Wolf Max Born. *Principles of Optics*. Cambridge University Press, Jan 1999.

[69] Yeong Min Jang, H Aoyama, R Roberts, Y Zeng, M Oshima, R Chitrakar, N Serafimovski, M Uysal, J Cha, SY Chang, et al. Technical considerations document. *online: https://mentor. ieee. org/802.15/dcn/15/15-15-0492-00-007a-technical-considerations-document. docx*, 2015.

[70] Roland Siegwart, Illah Reza Nourbakhsh, and Davide Scaramuzza. *Introduction to autonomous mobile robots*. MIT press, 2011.

[71] Richard Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[72] Ethan Rublee, Vincent Rabaud, Kurt Konolige, and Gary Bradski. Orb: An efficient alternative to sift or surf. In *2011 International conference on computer vision*, pages 2564–2571. Ieee, 2011.

[73] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[74] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3):346–359, 2008.

[75] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[76] Ivan Sipiran and Benjamin Bustos. Harris 3d: a robust extension of the harris operator for interest point detection on 3d meshes. *The Visual Computer*, 27(11):963–976, 2011.

[77] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. Brisk: Binary robust

invariant scalable keypoints. In *2011 International conference on computer vision*, pages 2548–2555. Ieee, 2011.

[78] David Nistér and Henrik Stewénius. Linear time maximally stable extremal regions. In *European Conference on Computer Vision*, pages 183–196. Springer, 2008.

[79] Jianbo Shi et al. Good features to track. In *1994 Proceedings of IEEE conference on computer vision and pattern recognition*, pages 593–600. IEEE, 1994.

[80] Pablo Fernández Alcantarilla, Adrien Bartoli, and Andrew J Davison. Kaze features. In *European conference on computer vision*, pages 214–227. Springer, 2012.

[81] Matlab. `https://www.mathworks.com/products/matlab.html`.

[82] Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In *Proceedings. international conference on image processing*, volume 1, pages I–I. IEEE, 2002.

[83] S. Sivaraman and M. M. Trivedi. Looking at vehicles on the road: A survey of vision-based vehicle detection, tracking, and behavior analysis. *IEEE Transactions on Intelligent Transportation Systems*, 14(4):1773–1795, 2013.

[84] Junhai Luo, Liying Fan, and Husheng Li. Indoor positioning systems based on visible light communication: State of the art. *IEEE Communications Surveys & Tutorials*, 19(4):2871–2893, 2017.

[85] Daniel Konings, Baden Parr, Fakhrul Alam, and Edmund M-K Lai. Falcon: Fused application of light based positioning coupled with onboard network localization. *Ieee Access*, 6:36155–36167, 2018.

[86] Zhice Yang, Zeyu Wang, Jiansong Zhang, Chenyu Huang, and Qian Zhang. Wearables can afford: Light-weight indoor positioning with visible light. In *Proceedings of the 13th annual international conference on mobile systems, applications, and services*, pages 317–330, 2015.

[87] Musa Furkan Keskin, Ahmet Dundar Sezer, and Sinan Gezici. Localization via visible light systems. *Proceedings of the IEEE*, 106(6):1063–1088, 2018.

[88] Shilin Zhu and Xinyu Zhang. Enabling high-precision visible light localization in today's buildings. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 96–108, 2017.

[89] Z. Zhao, P. Zheng, S. Xu, and X. Wu. Object detection with deep learning: A review. *IEEE Transactions on Neural Networks and Learning Systems*, 30(11):3212–3232, 2019.

[90] Xinchen Liu, Wu Liu, T. Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *ECCV*, 2016.

[91] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2167–2175, 2016.

[92] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David Anastasiu, and Jenq-Neng Hwang. Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*,

pages 8797–8806, 2019.

[93] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Lingyu Duan. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3235–3243, 2019.

[94] Clint Sebastian, Raffaele Imbriaco, Egor Bondarev, and Peter H. N. de With. Dual embedding expansion for vehicle re-identification, 2020.

[95] Bahador Khaleghi, Alaa Khamis, Fakhreddine O Karray, and Saiedeh N Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, 14(1):28–44, 2013.

[96] X. Chen, K. Kundu, Z. Zhang, H. Ma, S. Fidler, and R. Urtasun. Monocular 3d object detection for autonomous driving. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2147–2156, 2016.

[97] Bin Xu and Zhenzhong Chen. Multi-level fusion based 3d object detection from monocular images. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2345–2353, 2018.

[98] Xiaozhi Chen, Kaustav Kundu, Yukun Zhu, Huimin Ma, Sanja Fidler, and Raquel Urtasun. 3d object proposals using stereo imagery for accurate object class detection. *IEEE transactions on pattern analysis and machine intelligence*, 40(5):1259–1272, 2017.

[99] Khadija Ashraf, Vignesh Varadarajan, MD Rashed Rahman, Ryan Walden, and Ash-

win Ashok. See-through a vehicle: Augmenting road safety information using visual perception and camera communication in vehicles. *IEEE Transactions on Vehicular Technology*, 70(4):3071–3086, 2021.

[100] Ashwin Ashok, Khadija Ashraf, Vignesh Varadrajan, MD Rashed Rahman, and Ryan Walden. Non line-of-sight perception vehicular camera communication, 2021.

[101] Bin Yang, Wenjie Luo, and Raquel Urtasun. Pixor: Real-time 3d object detection from point clouds. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 7652–7660, 2018.

[102] Yin Zhou and Oncel Tuzel. Voxelnet: End-to-end learning for point cloud based 3d object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4490–4499, 2018.

[103] Ankith Manjunath, Ying Liu, Bernardo Henriques, and Armin Engstle. Radar based object detection and tracking for autonomous driving. In *2018 IEEE MTT-S International Conference on Microwaves for Intelligent Mobility (ICMIM)*, pages 1–4. IEEE, 2018.

[104] Carsten C Schwesig and Ivan Poupyrev. Radar-based object detection for vehicles, October 29 2019. US Patent 10,459,080.

[105] Charles R Qi, Wei Liu, Chenxia Wu, Hao Su, and Leonidas J Guibas. Frustum pointnets for 3d object detection from rgb-d data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 918–927, 2018.

[106] Xiaozhi Chen, Huimin Ma, Ji Wan, Bo Li, and Tian Xia. Multi-view 3d object de-

tection network for autonomous driving. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1907–1915, 2017.

[107] Ming Liang, Bin Yang, Shenlong Wang, and Raquel Urtasun. Deep continuous fusion for multi-sensor 3d object detection. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 641–656, 2018.

[108] Jason Ku, Melissa Mozifian, Jungwook Lee, Ali Harakeh, and Steven L Waslander. Joint 3d proposal generation and object detection from view aggregation. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–8. IEEE, 2018.

[109] Charles Ruizhongtai Qi, Li Yi, Hao Su, and Leonidas J Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space. In *Advances in neural information processing systems*, pages 5099–5108, 2017.

[110] Shenlong Wang, Simon Suo, Wei-Chiu Ma, Andrei Pokrovsky, and Raquel Urtasun. Deep parametric continuous convolutional neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2589–2597, 2018.

[111] Ming Liang, Bin Yang, Yun Chen, Rui Hu, and Raquel Urtasun. Multi-task multi-sensor fusion for 3d object detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.

[112] Piotr Dollár, Ron Appel, Serge Belongie, and Pietro Perona. Fast feature pyramids for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(8):1532–1545, 2014.

[113] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *CoRR*, abs/1804.02767, 2018.

[114] Tzutalin. labelimg. git code (2015). `https://github.com/tzutalin/labelImg`.

[115] AbdulHaseeb Ahmed, Sethuraman Trichy Viswanathan, MD Rahman, and Ashwin Ashok. An empirical study of deep learning models for led signal demodulation in optical camera communication. *Network*, 1(3):261–278, 2021.