DESIGN, MODELING, AND ANALYSIS OF VISUAL MIMO COMMUNICATION

By

ASHWIN ASHOK

A Dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

written under the direction of

Dr. Marco O. Gruteser, Dr. Narayan B. Mandayam and Dr. Kristin J. Dana

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

October, 2014

ABSTRACT OF THE DISSERTATION

Design, Modeling, and Analysis of Visual MIMO Communication

By ASHWIN ASHOK

Dissertation Director:
Dr. Marco O. Gruteser, Dr. Narayan B. Mandayam and Dr. Kristin J. Dana

Today's pervasive devices are increasingly being integrated with light emitting diode (LED) arrays, that serve the dual purpose of illumination and signage, and photo-receptor arrays in the form of pixel elements in a camera. The ubiquitous use of light emitting arrays (LEA) and cameras in today's world calls for building novel systems and applications where such light emitting arrays can communicate information to cameras. This thesis presents the design, modeling and analysis of a novel concept called *visual MIMO* (multiple-input multiple-output) where cameras are used for communication. In visual MIMO, information transmitted from light emitting arrays are received through the optical wireless channel and decoded by a camera receiver. The paradigm shift in visual MIMO is the use of digital image analysis and computer vision techniques to aid in the demodulation of information, contrary to the direct processing of electrical signals as in traditional radio-frequency (RF) communication.

The unique aspect of camera communications is that visual perspective distortions dominate over distance-based attenuation, multipath fading and other important properties of the radio-frequency (RF) wireless channels. In visual MIMO, camera receivers together with LEAs allow multiple parallel channels as in RF MIMO to achieve throughput gains, but these gains depend on the perspective—orientation and distance—between the transmitter and receiver. Camera receivers also allow for a large field-of-view for signal reception and can facilitate tolerating mobility through intelligent tracking techniques to locate the light emitting transmitter in view. This dissertation studies these key aspects of visual MIMO communication, and has been structured into three parts. The first part derives the perspective dependent channel model and information capacity of visual MIMO communication, along with a case-study of capacity of camera communication using display screens as transmitters. Part two discusses perspective dependent throughput enhancement techniques that exploit the MIMO array structure and uses vehicle-vehicle (V2V) communication as a running example. Finally, part three discusses transmitter localization techniques that help adapt to mobility in visual MIMO channels. The inferences and lessons learned through this thesis open up novel opportunities to use cameras as an integral part of a communication system. Efforts have already been initiated by the optical wireless communication community to standardize camera communications, and such advances attest to the importance of using cameras for communications.

Acknowledgements

I would like to thank my advisors Profs. Marco Gruteser, Narayan Mandayam and Kristin Dana who played a pivotal role in training me through my tenure as a Ph.D. student. I appreciate their unwavering support and advices which helped me shape up myself as a passionate researcher and my thesis as well. More importantly, they made my journey intellectually very enlightening and lot of fun. They still continue to motivate and inspire me. Their passion in research and their humility and personal support is something I will always keep close to my heart and mind, and will strive to imbibe in myself as I progress in my career.

Next, I would thank Dr. Richard Howard, whom I refer to as a walking encyclopedia of knowledge and wisdom. Every time I have interacted with him I end up learning twice as much. He is a great inspiration to me as "the real scientist" and the passion in him for science is amazing. I find him in my list of academic heroes. I thank him whole-heartedly for the wonderful collaboration I had through my research projects and as a mentor at WINLAB.

I would like to thank WINLAB as a whole, including all the faculty and staff, who make WINLAB a wonderful place to pursue research. I had the perfect ambiance around me at WINLAB and I would like to extend my respect to WINLAB. I would like to specially thank Prof. Dipankar Raychaudhuri, Director of WINLAB, for his belief and support in the visual MIMO project. I want to thank Ivan Seskar for his unwavering support and advice in my experiments at WINLAB. His knowledge and passion in research was one of the key inspirations for me to consider WINLAB. I would like to also thank Prof. Roy Yates, with whom I worked as a teaching assistant for two wonderful semesters. He is a great inspiration and example to me as the right blend of a great teacher and researcher. I would also like to thank Dr. Yanyong Zhang

whom I collaborated with and who inspired and taught me many ways of tackling tough research problems.

I would like to thank National Science Foundation for their monetary support of the visual MIMO project for the tenure of my Ph.D., and all my collaborators in and out of WINLAB who have helped me shape my thesis work. I would like to thank Prof. Thomas Little, from Boston University, for serving in my committee and guiding me through my research. His remarks and comments played a pivotal role in shaping my thesis.

Last but not the least, I would like to thank all my group-mates, colleagues, friends and family members for the support and love throughout my Ph.D. student tenure. The days I have spent in WINLAB and Rutgers will be some of the most memorable days in my life and the list of people whom I would like to thank is never-ending. With all due respect, I would like to express my earnest gratitude to my parents and my family members who have been supporting me always, with love, care and respect. Thanks to all of you!

Dedication

To Mom, Dad, Sister, and a few special people in my life. This thesis is a dedication to all of them.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

We live in a world where we are ever more surrounded by communications and entertainment devices, and the convergence of communications, information and entertainment has never been more apparent. Today's pervasive devices are increasingly being integrated with light emitting elements in the form of light emitting diode (LED) arrays, that serve the dual purpose of illumination and signage, and photo-receptor arrays in the form of pixel elements in a camera.

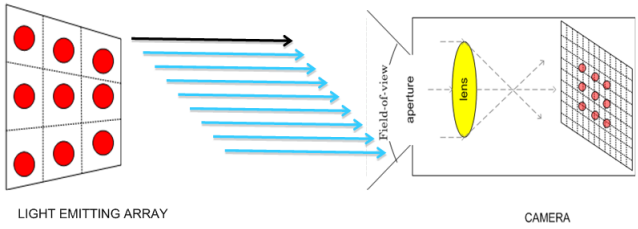


Figure 1.1: Illustration of visual MIMO concept. A light emitting array (LEA) communicates information to a camera

The increasingly ubiquitous use of cameras (or photodiode arrays) in pervasive devices such as smartphones and tablets, cars, gaming consoles, etc. creates an exciting opportunity to use cameras for more than just photo/video capturing. The ubiquitous use of light emitting arrays (LEA) and cameras in today's world calls for building novel systems and applications where such light emitting arrays can communicate information to cameras. This thesis presents a novel concept in this direction called *visual MIMO* (multiple-input multiple-output) that will enable cameras to be used for communication. In visual MIMO, optical transmissions from light emitting arrays, are received by an array of photo diode elements (e.g. pixels in a CMOS camera). Examples of LEAs

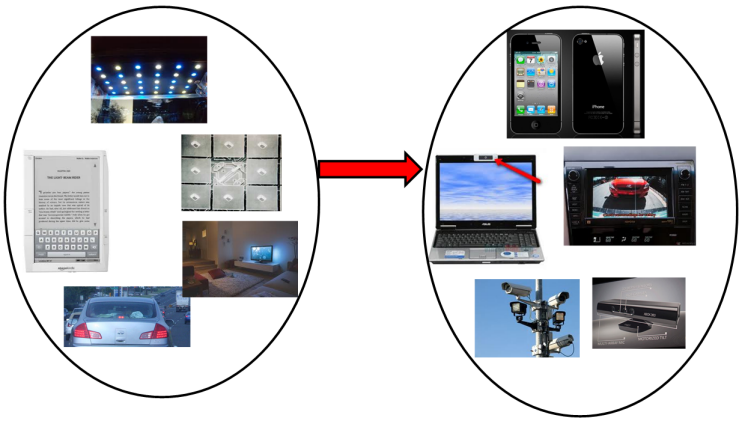


Figure 1.2: Illustration of possible off-the-shelf LEA transmitters and camera receivers in visual MIMO communication applications

include arrays of light emitting diodes (LED), pixel arrays on LCD or plasma screens, or digital micro mirror devices combined with a light source such as in projectors. In addition, the array of LEDs in lighting arrays and commercial display devices, LCD pixels in display screen, projector screens, and even printed material also qualify as potential transmitters in visual MIMO. Figure 1.1 provides a conceptual illustration of the visual MIMO concept.

In visual MIMO, information is transmitted using light emitting arrays that are detected and decoded by the photo receptor elements or pixels of a camera receiver. In principle, visual MIMO is unique when compared to traditional optical wireless or free-space optics. Visual MIMO uses a camera as a receiver instead of custom built photo receptor receivers. Visual MIMO leverages the inherent 2D spatial array structure of the image sensor pixels to create a multi-input-multi-output (MIMO) channel. The MIMO analogy stems from the fact that visual MIMO treats each LED element of the LEA as a transmit antenna and each pixel of the camera as a receiving (photodetector) antenna. The key motive of this thesis is to develop the visual MIMO concept to design and implement novel optical wireless communication systems and applications that can use cameras (integrated or custom) for communication.

1.2 Applications

Several key applications in diverse fields could benefit from visual MIMO. Figure 1.2 shows some of the possible off-the-shelf LEAs transmitters and camera receivers that can apply the visual MIMO concept.

Vehicular communications. Safety applications in vehicular networks such as emergency electronic brake lights [2] and cooperative collision warning (CCW) [3] require reliable communications under potentially high co-channel interference because vehicle position and dynamics information needs to be shared among nearby vehicles in potentially very dense highway scenarios. Visual MIMO could potentially reuse emerging automotive LED lights and cameras to create an alternate communication channel that reduces interference because its directional and line-of-sight transmissions allow for increased spatial reuse.

Pervasive mobile communications. The ubiquitous placement of electronic screens and surveillance cameras in urban environments create numerous opportunities for practical applications of visual MIMO channels. Screens for electronic signage can have dual functionality by transmitting embedded signals so that visual observation for human observers would coexist with a visual MIMO wireless communications channel. This could enable novel advertising applications; imagine a smartphone user pointing the handset camera at one of the numerous electronic billboards to receive further information such as a purchase URL. Signals might be embedded through intensity modulation or angle-based modulation where observation of the screen at different angles enables different visual observation. Such embedded signals may also enable new human-computer user interfaces, for example by facilitating recognition of pointing or gestures with a camera-equipped mobile device.

Unobtrusive and 'secure' communications. Tactical communications can benefit from the security properties of visual mimo channels. The line-of-sight requirement greatly reduces the potential for interception and jamming that is inherent in RF communication. Additionally, the source of the signal can be more easily determined, so

the potential for spoofing signals is reduced. Furthermore, visual MIMO receivers attached to surveillance cameras and transmitters embedded in cell phones could provide a backup communication channel for first responders or facilitate localization of 911 callers. For example, when a person dials 911 in a large inner-city building where current E911 localization is not sufficiently precise, the emergency operator could ask the person to use the phone as a visual beacon and point the screen towards a nearby surveillance camera. The camera could detect this signal and might send this signal together with the camera location to the emergency dispatch center, or simply call for attention from a human camera operator.

Environment mapping, localization and recognition. Visual MIMO also finds application in computer vision, where camera networks refer to the cooperation of numerous cameras viewing a scene in order to create a 3D environment map, which can benefit localization and recognition. An interesting merger of computer vision recognition algorithms with communications protocols can be explored by recognizing not static passive objects, but objects that are communicating known temporal pilot sequences and headers. Cameras can facilitate very precise positioning and localization through image analysis and projection theory from computer vision. While off-the-shelf cameras already provide the necessary tool-kit for this purpose, designing custom cameras and camera-like devices offer more flexibility in terms of optimization in the parameter space such as size, power or efficiency.

Applications and prototypes from this thesis work. As will be discussed in the following chapters, this thesis also identifies key applications of visual MIMO and implements prototypes for the same. In particular, this thesis uses three use-case application scenarios and develops working prototypes for the same:

- *Vehicle-to-Vehicle communication*, where brake-lights (or head-lights) of cars and vehicles can be used for communicating to cameras fit in vehicles. One example use-case is where the vehicle's sensor data, such as speed of the vehicle, is transmitted to the vehicles that follow on the road.

- *Pervasive mobile applications*, where cameras in smartphones and other pervasive

mobile devices can be used to retrieve information from ubiquitous light transmitters such as LED and LCD displays. One example use-case is to use such display screens to phone-camera interactions for ticketing and advertising.

- *Positioning and wearable computing*, where cameras can be built as wearable devices that help accurately recognizing a object's identity and location in space. The cameras can be customized for low-power battery operation (battery operated wearable devices) and can also be integrated with existing technology such as radio for energy conservation as well as accuracy.

The prototypes developed in this thesis aim at a proof-of-concept demonstration of visual MIMO systems. In general, this thesis work aims to set a baseline model and prototype designs for future camera based communication applications that will use the visual MIMO concept.

1.3 Unique aspects of Visual MIMO

Visual MIMO takes advantage of existing cameras and light emitting devices, and it possesses several distinct properties that can present advantages over radio-frequency (RF) based wireless communications and conventional free-space optics. These properties also define the unique characteristics of visual MIMO communication:

- **Directional and long-range communication.** The image sensor in a camera is essentially an array of photodiodes and the camera lens provides a different narrow field of view for each photodiode. This creates a large number of highly directional receive elements (the camera pixels), which allows reducing interference and noise and thereby can achieve large ranges. The use of cameras and LEAs for communication can help overcome the transmission range limitations of conventional wireless optics.

- **Uncorrelated channels enabling spatial multiplexing.** In visual MIMO, the system can transmit with multiple LEDs and record the signal with multiple camera pixels. This approach can also allow many "parallel" or uncorrelated

communication channels, similar in concept to RF MIMO systems [4], albeit over a channel with very different characteristics. Uncorrelated channels allow for multiplexing bits over each channel and decode all at camera receiver at the same time (in parallel). It also allows for multiple access where the camera can receive signals from multiple light emitting devices at the same time. The inherent structure of visual MIMO channels allows for much easier spatial multiplexing as compared to its radio frequency counterpart.

- **Wide field-of-view enabling mobility.** The image sensor in a camera is essentially an array of photodiodes and the camera lens provides a different narrow field of view for each photodiode. The lens in cameras directs the light components separately to each pixel, thus enhancing the overall field-of-view of the lens + photodiode array system; the camera field-of-view is the sum of field-of-view of each pixel. This structure creates a large number of highly directional receive elements (the camera pixels), which allows reducing interference and noise and thereby can achieve large ranges and tolerating mobility.

1.3.1 Challenges in camera communications

Camera limitations. Visual MIMO allows for multiple spatial parallel channels and thus large-data rate communication is possible by using arrays with large number of light emitters. The trade-offs in the visual MIMO system, however, are a limited receiver sampling frequency or frame-rates; sampling frequency is equivalent to (temporal) bandwidth of communication. Frame-rates of cameras today are of the order of hundreds to thousand frames per second for lower end cameras and a million frames per second for high-end models. However, commercial cameras, such as video cameras and mobile cameras, are limited to 30-60fps; and 120fps slow-motion cameras such as in Iphone 5S. Moreover, cameras today are also limited in quantization due to the digitization process. Camera pixel intensities are usually 8 bit (in each channel) which limits the number of quantization levels of the light beam. One possible solution is to use sophisticated cameras with better quantization and sampling rates, but the

trade-off is the high cost and complexity. Another solution is to build custom cameras for communication. Cameras are also power hungry devices and operating cameras in an always-ON mode can drain a lot of power. Depending on the application power consumption will be a key factor in a camera design for communication; for example, battery powered cameras in mobile devices. Customization can offer a higher level of flexibility and control of the camera (and hence the receiver) specifications. However, such customizations will require the additional effort to build such cameras that can also cater to diverse applications.

Perspective distortions. In visual channels, perspective distortions dominate over some of the important properties of a RF wireless channel such as distance based attenuation, multipath fading and doppler shifts. In visual MIMO, the received signal at any instance is an image of the transmitting element along with the background scene or imagery. Perspective distortions in camera channels manifest as reduction in the size of the imaged light emitting element, with distance and skew/rotation in the image due to angular view. In addition, lens blur (typically due to focus imperfection or jerks while capturing the image) additionally affect the image quality. In an ideal scenario (pin-hole camera), each light emitter has a dedicated "channel" to each pixel of the camera. In this scenario all the channels in visual MIMO are uncorrelated. However, in reality, the camera lens creates the perspective effect where the light rays are directed to a finite space (image sensor) and thus the number of such uncorrelated channels depend on the placement of the object/scene with respect to the lens. For example, when the object is placed far away from the lens the object looks much smaller than it was when placed nearer. A similar effect will be observed when the object is placed at an angle from the lens since the light rays will have to travel a longer distance as compared to the fully frontal view.

1.4 Thesis Objectives

In particular, the objective of this thesis to design a visual MIMO communication by identifying it's unique characteristics and challenges. This thesis is motivated by the following hypotheses:

1. **Perspective dependent channel.** The information capacity of visual MIMO is largely dependent on the perspective between the camera and the light emitter. A model that accounts for such perspective dependency will be beneficial.

2. **Throughput gains by adapting to visual perspectives.** It is possible to achieve large throughput (data-rate) gains in visual MIMO by leveraging the spatially uncorrelated channels and techniques that help adapt to the visual channel distortions. However, these gains depend on factors such as visual distortions, geometry and size of transmitter arrays, as well as the camera specifications.

3. **Visual MIMO for low-power localization.** Cameras have a wide field-of-view that enables mobility in visual channels by locating the transmitting light element in the camera sampled image. Cameras are power hungry and not efficient for an always-ON operation as required by visual localization system that typically use mobile tracking and recognition. The directionality and wide field-of-view of array receivers (as in visual MIMO) can be leveraged to build custom low-power camera-like receiver devices that facilitates tracking, positioning and localization.

The following sections will elaborate on these points and discuss the specific contributions of this thesis.

1.4.1 Perspective dependent channel

A camera channel is analogous to a RF MIMO channel where each pixel element of the camera acts as a receiving antenna and the light emitting elements as the transmit antennas. In RF MIMO, the signal quality at each receive antenna element is a function of the path-loss in the channel, multipath fading, and the interference from other transmit antennas — also called co-channel interference [4]. A camera channel has negligible multipath fading but experiences path-loss in light energy, and interference (of light energy) from other light emitting elements, which manifest as distortions in size and shape of the imaged light emitter on the output of a camera. The paradigm shift in visual MIMO is the use of the camera's pixel array structure to facilitate the use of image analysis and computer vision techniques to aid in the demodulation of information,

contrary to the direct processing of electrical signals at the receiver as in traditional radio-frequency (RF) communication. However, the performance of camera receivers depend on some unique factors such as perspective distortions, quantization in (digital) image sensors, lens artifacts such as blur, frame-rate of cameras and synchronization between light emitter and cameras.

Contribution This thesis develops a communication model for visual MIMO that captures the perspective dependence of the visual channel. Unlike radio channels where distortions are random, visual distortions are very geometry dependent. Our model uses classical camera imaging theory and incorporates projection theory from the computer vision domain. The model is used to study and derive the information capacity of visual MIMO communication. Our capacity model accounts for perspective dependent (position and orientation) distortions that dominate this channel and other factors unique to cameras and light emitting devices. The lessons learned from studying the visual MIMO model is used as a baseline to derive and evaluate the capacity of a use-case visual MIMO application where the transmitter is a display screen.

1.4.2 Throughput gains by adapting to visual perspectives

The visual MIMO channel allows highly directional transmission and reception, thus attractive for very dense congested environments. The array structure in cameras can allow for a large number of spatially uncorrelated channels rendering the communication virtually interference-free. Information can be multiplexed over these spatially uncorrelated channels almost seamlessly. Due to negligible multipath fading in optical channels the data-rates achievable (i.e., the degree of multiplexing) in visual MIMO depend primarily on the distortions in the visual channel rather than multipath fading as in RF. In mobile settings, the quality of the visual MIMO link varies significantly with the variation in these distortions which depends on the camera receiver perspective. An ideal pin-hole model would essentially yield uncorrelated channels. Due to the nature of the lens in the camera the degree of correlation/uncorrelation depends on perspective. For example, at long distance most of the channels are correlated – depending on the size of the LED array.

Contribution To improve throughput of visual MIMO links this thesis develops rate adaptation techniques which adapt the transmission data rate to the receiver perspective. The rate adaptation challenge in visual MIMO lies primarily in MIMO *mode* adaptation. The modes are defined as the logical set spatial combinations of transmitting elements chosen by the transmitter when communicating with the camera. These modes present a more complex set of choices than what RF rate adaptation algorithms have explored. The rate adaptation problem then is to choose transmission modes that exploit the available parallel channels while keeping the error rate low. Multiplexing across more transmitter elements will lead to higher data rates, but including an LED that is occluded in the image, for example, would lead to bit errors.

1.4.3 Visual MIMO for low-power localization

Optical wireless communications with narrow beams has hitherto been impractical in most mobile settings, because both the sender and receiver need to operate with very narrow beams and angles-of-view, respectively, to achieve transmission ranges greater than a few tens of meters. Due to the extremely narrow beam-widths used, any application with some mobility would require costly mechanical steering systems for transmitter and receiver. This problem can be alleviated using visual MIMO approach through techniques that help acquire and track signals from a transmitter as they are captured by different photo-receptor elements during movement. However, cameras are typically power hungry sources and so alternate low-power solutions may be required.

Contribution This thesis develops a hybrid radio-optical approach that enables mobility in visual channels through high accuracy angle-of-arrival and ranging based positioning. This approach leverages the high directionality characteristic of the optical link for precise angle-of-arrival estimation and ranging. A low-power radio link is used to communicate ID as well as synchronize the optical and radio transmissions, thus conserving energy. The prototype design uses a geometrical arrangement of few photodiodes to build a camera-like receiver, to estimate the angle-of-arrival and distance between a radio-optical (LED) transmitter and receiver based on the sampled optical signal strength on these photodiodes. The identities and synchronization control is

achieved through the radio channel using radio-frequency IDs (RFID).

1.5 Thesis Organization

This thesis is organized as follows: The current chapter introduces the visual MIMO concept, that forms the key element of this thesis, and outlines the thesis contributions. Chapter 2 derives the visual MIMO channel model and information capacity of visual MIMO communication followed by a case-study of capacity of camera communication using display screens as transmitters in Chapter 3. Chapter 4 will discuss the visual MIMO rate adaptation techniques to enhance throughput in visual MIMO links. Chapter 5 will discuss a radio-optical approach that helps design a low-power visual MIMO positioning system. Elaborating on future directions of this thesis, Chapter 6 will conclude this thesis. This dissertation thesis address three key aspects of a system design, (i) channel modeling and analysis, (ii) algorithm and hardware design, and (iii) prototyping and experimentation.

Chapter 2

Visual MIMO Communication System Model

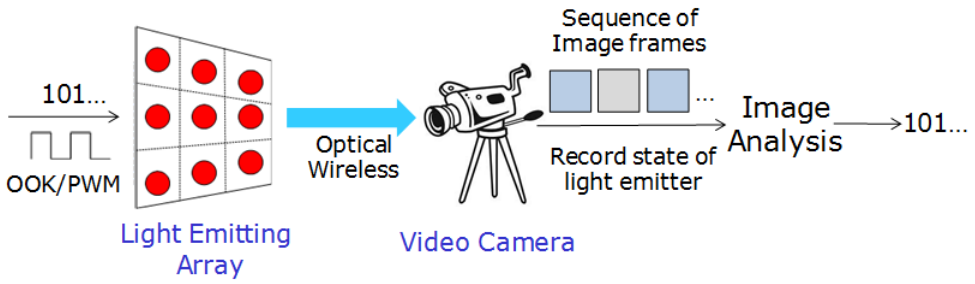


Figure 2.1: Illustration of visual MIMO system functionality. Information is modulated as ON-OFF or intensity patterns of light emitted from light arrays and received and processed by a camera (OOK = ON-OFF Keying, PWM = Pulse Width Modulation)

.

Advances in CMOS imaging technology along with the advent of visible and infrared (IR) light sources such as light emitting diode (LED) arrays or LCD screens present an exciting and challenging opportunity to enable *mobile* optical networking through a novel concept developed in this thesis, called *visual MIMO*. In this concept, optical transmissions by multiple transmitter elements are received by an array of photodiode elements (e.g. pixels in a CMOS camera). The paradigm shift in this design is the use of image analysis and computer vision techniques at the receiver. Figure 2.1 shows a functional illustration of the visual MIMO concept.

In the visual MIMO communications, the optical transmit element generates a light beam (optical signal) whose output power is proportional to the electrical input power of the modulating signal, limited by the emitter's peak transmission power. While RF channels are typically characterized by their impulse response that reflects the multipath environment, this aspect differs significantly for optical channels. Since the rate of change of the channel impulse response is very slow compared to the frequency of

the optical signal, it is usually sufficient to use a static parameter (channel DC gain) [5] to represent the channel. For the same reason inter-symbol interference and multipath fading can be neglected in optical wireless channels. Similarly Doppler shift is negligible since optical frequencies (order of THz) are much higher compared to radio.

In this chapter we develop a channel model for visual MIMO communication, and study and evaluate its key properties.

2.1 Visual MIMO Channel Model

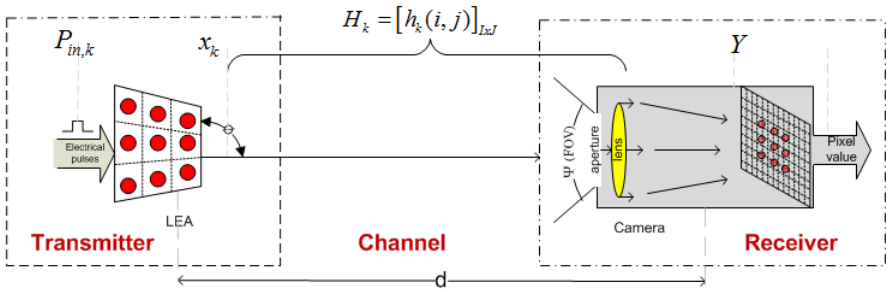


Figure 2.2: The LEA-Camera visual MIMO communication model

Let us consider an optical transmitter consisting of an array of K transmitting elements communicating to a photodiode array (camera) receiver with $I \times J$ elements (pixels). We will refer to the 2D array of signal values obtained from the 2D transmit/receive elements in visual MIMO communication as an 'image'. The channel model for the visual MIMO system as shown in Fig. 2.2 can be expressed as,

$$\mathbf{Y} = \sum_{k=1}^{K} \mathbf{H}_k x_k + \mathbf{N} \tag{2.1}$$

where $\mathbf{Y} \in \mathcal{R}^{I \times J}$ is the image current matrix with each element representing the received current $y(i,j)$ in each pixel with image coordinates (i,j), $x_k \in \mathcal{R}$ represents the transmitted optical power from k^{th} element of the LEA and $\mathbf{H}_k \in \mathcal{R}^{I \times J}$ is the channel matrix of the k^{th} transmit element of the LEA, with elements $h_k(i,j)$ representing the channel between the k^{th} transmit element and pixel (i,j), and $\mathbf{N}$ is the noise matrix. Noise in optical wireless is dominated by shot noise from background light sources

and typically modeled as AWGN [5, 6]. Each element $n(i,j)$ of the noise matrix $\mathbf{N}$ representing the shot-noise current at each receive image pixel is given as,

$$n(i,j) = \sqrt{\sigma_{shot}} = \sqrt{2qRP_n s^2 W} \tag{2.2}$$

where q is the electron charge, R is the responsitivity of the receiver characterized as the optical power to current conversion factor, P_n is the background shot noise power per unit area, s is the square pixel side length and W is the sampling rate of the receiver (equates to the frame rate of the camera).

In visual MIMO communication the optical signal from the k^{th} transmit element ($k = 1, 2, 3 \ldots K$) emitting a light beam of power $P_{in,k}$ will be transmitted over free-space. At the receiver, depending on the focusing of the camera and the distance between the transmitting element and the camera, the transmitting element's image may strike a pixel or a group of pixels of the detector array. The signal current in each pixel will depend on the concentration of the received signal component on that pixel which can be quantified as the ratio of the pixel area relative to the area spanned by the transmitting element's image on the detector. If $c_k(i,j)$ represents the concentration ratio of the k^{th} transmit element of an LEA on pixel (i,j), the channel DC gain $h_k(i,j)$ from each transmit element k to the pixel (i,j) is given as

$$h_k(i,j) = R \times R_o(\Phi) \times A_{lens} \times \frac{\cos(\psi)\cos^2(\phi_{k,i,j})}{d^2_{k,i,j}} \times c_k(i,j) \tag{2.3}$$

where $R_o(\Phi)$ is the Lambertian radiation pattern of the optical transmitting element [5] with half-power angle Φ, A_{lens} is the area of the camera lens, ψ is the camera field-of-view (fov) and $d_{k,i,j}$, $\phi_{k,i,j}$ are the distance & viewing angle between each transmit element k and receiving pixel (i,j) respectively.

Typically, since the pixel size is very small (order of microns), the difference in distance $d_{k,i,j}$ and the viewing angle $\phi_{k,i,j}$ between each element of the transmitter array and every pixel is negligible. Therefore we refer to the distance $d_{k,i,j} = d$ and the viewing angle $\phi_{k,i,j} = \phi$ as the perpendicular distance and the angle between the transmitter array and image detector planes respectively. Hence the channel between

each transmit element k and each pixel (i, j), characterized by $h_k(i, j)$, is primarily dependent on the concentration ratio,

$$c_k(i,j) = \frac{s^2}{\pi(\frac{fl_k}{d} + \sigma_{blur})^2/4}\mathcal{I}_k(i,j) \tag{2.4}$$

$$\mathcal{I}_k(i,j) = \begin{cases} 1 & \forall(i - i_k^{ref})^2 + (j - j_k^{ref})^2 \leq (\frac{fl_k}{d} + \sigma_{blur})^2/4 \\ 0 & otherwise \end{cases} \tag{2.5}$$

where, s, f, l_k are the pixel edge length, camera focal length and diameter of k^{th} transmit element (considering a circular transmitting element) respectively. The amount of concentration of the signal per pixel is also dependent on the amount of blur in the image due to the lens. Typically, lens blur is modeled as a Gaussian function [7] and the amount of blur in the image is quantified by its standard deviation (σ_{blur}). The lens essentially acts like a filter with the blur function as its impulse response. Thus the image of the transmit element can be viewed as a result of the projected image convolving with the blur function over the detector area. $\mathcal{I}(.)$ is an indicator function indicating whether a pixel (i, j) receives a signal from the transmit element k or not, and is referenced in terms of the distance from pixel at the center of the transmit element's image (i_k^{ref}, j_k^{ref}). Given the spatial coordinates of the transmitting elements of an LEA with respect to the camera reference we can determine the image center coordinates of those transmit element through optical ray-tracing techniques in conjunction with projection theory from computer vision [8].

The optical noise on each camera image pixel comprises of the ambient lighting from the environment and any amplifier noise in the imaging circuitry. This noise is typically signal independent when the ambient lighting is sufficiently high, such as in office rooms, sunlight [9] and are typically characterized as additive-white-Gaussian-noise (AWGN). In a camera image pixel, this noise is quantized and manifested as fluctuations in the gray-level pixel intensity, which is the digital value of the sensor output. Noise from background lighting can be considered isotropic over the receiver surface area due to the small size of the image sensors (or photodiode arrays) and quantified through the

AWGN noise-variance σ_n^2; the noise power per pixel averaged over the entire image sensor area [9].

2.2 Perspective Dependent MIMO Channel

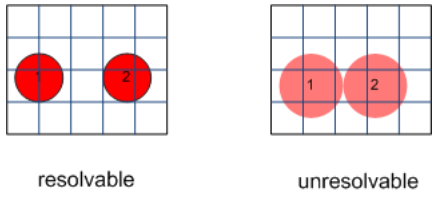


Figure 2.3: Effect of lens blur on the resolvability of images

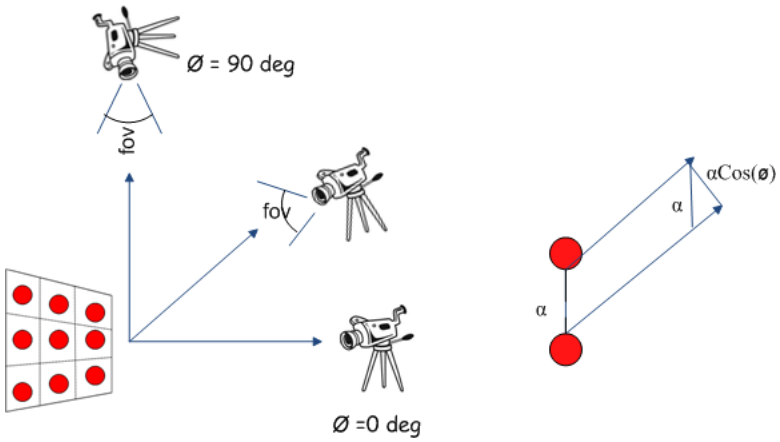


Figure 2.4: Camera Viewing angle Illustration

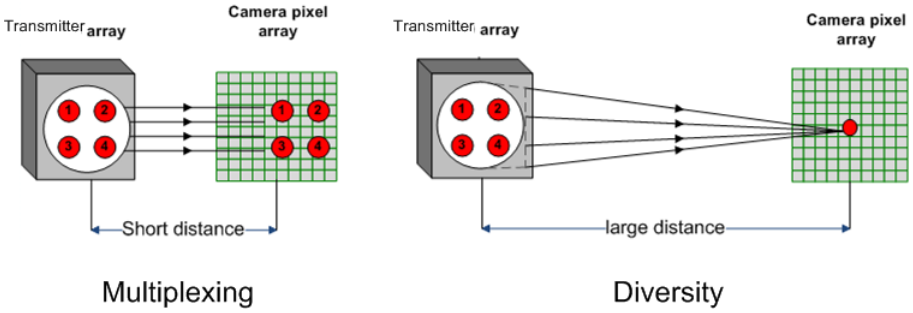


Figure 2.5: Distance dependent Multiplexing and Diversity modes

In visual MIMO, the system can transmit with multiple light elements and record the signal with multiple camera pixels. This approach can also allow many "parallel" or uncorrelated communication channels, similar in concept to RF MIMO systems [4], albeit over a channel with very different characteristics. Uncorrelated channels allow

for multiplexing bits over each channel and decode all at camera receiver at the same time (in parallel). It also allows for mutiple access where the camera can receive signals from multiple light emitting arrays at the same time. The inherent structure of visual MIMO channels allows for much easier spatial multiplexing as compared to its radio frequency counterpart.

In principle, a camera is essentially an array of photoreceptors (on an image sensor) with a lens fit in front. Photoreceptor arrays along with lenses undergo perspective distortion, which is a common effect seen in camera images where the image object undergoes deformation in size and shape. If multiple light emitting elements are spaced very close to each other then the resultant effect of such deformations lead to multiple light rays from the same object interfering on the same pixel. We term this effect what we call as 'interpixel interference (IPI)'and discuss in more detail in Chapter 3. In this section, we will study the MIMO characteristics of visual MIMO communication. Unlike fading in RF wireless channels, these distortions are deterministic and are caused due to the nature of camera imaging process. While the channel model in (2.1) resembles that of the familiar RF MIMO channel model, in fact it is significantly different from that. In RF MIMO systems, the channel matrix is typically a rich scattering matrix (usually full rank) whose entries are modeled well as independent and identically distributed random variables [10]. Further, this property allows the RF MIMO system to exploit either diversity and or multiplexing gains in data transmission which primarily depend on the multipath fading in the RF channel. The fact that the communication system here uses light as the communication medium, requires line of sight at the receiver, and the nature of the concentration function of the camera, renders some unique characterizations different from RF MIMO.

The notion of 'parallel'channels to obtain the multiplexing data rate gains can be achieved only if the circumference (assuming circular light emitting transmit elements) of two transmit elements as seen on the image plane are separated by atleast a threshold (η) number of pixels in both dimensions (horizontal and vertical). As we see in Fig. 2.3 even if the circumference of the two transmit element images are separated by one pixel they may not be resolvable because of the blur in the image. Hence we set a threshold

distance of separation between the image circumferences, $\gamma = 2\sqrt{2ln2}\sigma_{blur}$, equal to the full-width-half maximum (FWHM) of the Gaussian lens-blur function typically used as a parameter for image resolution in analyzing fine detailed astronomical and medical images [11,12]. The distance of separation between the images of the transmit elements can be determined by perspective projection analysis (as described in [13]) considering circular transmitting elements. Given a fixed-focal length f of the camera, pixel side length s and a spatial distance α between the circumference of two adjacent LEDs, the circumference of two transmit element images will be separated by $\alpha_{im} = \frac{f\alpha}{ds}$ pixels in each dimension. Therefore the separation between the circumference of two transmit element images will be equal to the threshold (γ/s) at a distance $d^* = \frac{f\alpha}{\gamma}$ between the LEA and camera. This implies that, multiplexing in visual MIMO is possible only when $d \leq d^*$ and when $d > d^*$ each transmit element has to transmit the same information whereby diversity combining at the receiver can ensure an SNR gain and hence an equivalent capacity gain.

We can observe in equation (2.3) that the channel quality depreciates with the viewing angle ϕ (angle between the camera image plane and LEA surface plane). Two images which are clearly separated in the image plane may look overlapped when viewed from an angle. Such distortions can significantly depreciate the signal quality and the detection capability leading to errors and thus reduction in the data rates. Moreover such an angular view also reduces the achievable multiplexing transmission range. This is because when the camera image detector plane is at an angle ϕ to the transmitter array the effective spatial separation between two neighboring transmit elements becomes $\alpha cos(\phi)$ $(\leq \alpha)$ (as shown in Fig. 2.4). From the earlier discussion on the resolvability of images, it implies that the distance upto which multiplexing can be achieved in visual MIMO then reduces to

$$d^* = \frac{f\alpha}{\gamma} cos(\phi) \tag{2.6}$$

In the visual MIMO channel, for a static transmitter and receiver, the image of the LEA transmit elements captured by the camera may span one pixel or multiple pixels. Further, the image plane is spanned by images of each transmit element clearly

delineated and the size of image span depending on the focus (concentration ratio) of the camera. As illustrated in Fig. 2.5, at short distances between the transmitter and receiver, each transmitting element of the LEA looks clearly focused on a unique set of pixels and the images of these elements can be detected from the complete image. In contrast, at a large distance between the transmitter and receiver, the image of each transmit element looks clearly unfocused and thus the signal from all the transmitting elements of the LEA is directed to typically one or few pixels. This suggests that at short distances, the system can offer large "multiplexing" gains by using the transmitting elements to signal independent bit-streams or equivalently realizing parallel channels. On the other hand, at large distances, there can only be a "diversity" gain where by the same bits are signaled on each of the transmit elements. These distance dependent gains in visual MIMO is in contrast to the RF MIMO channel, where the rich scattering channel matrix typically allows a continuous trade-off between diversity and multiplexing gains [14, 15].

2.3 Channel Capacity Analysis

Considering the AWGN channel and deterministic nature of perspective distortions, capacity (measured in bits/sec) of visual MIMO communication can be expressed using Shannon Capacity formula as,

$$C = \frac{W_{fps}}{2} W_s log_2(1 + SINR) \tag{2.7}$$

where $SINR$ represents the average signal-to-interference noise ratio per pixel, W_{fps} is temporal bandwidth or frame sampling rate (frames-per-second) where the factor 2 in the denominator corresponds to the Nyquist sampling rule, and W_s is spatial-bandwidth or the number of information carrying photoreceptor elements at the receiver image sample (pixels per camera image frame).

We will now consider a case-by-case analysis of visual MIMO channel capacity.

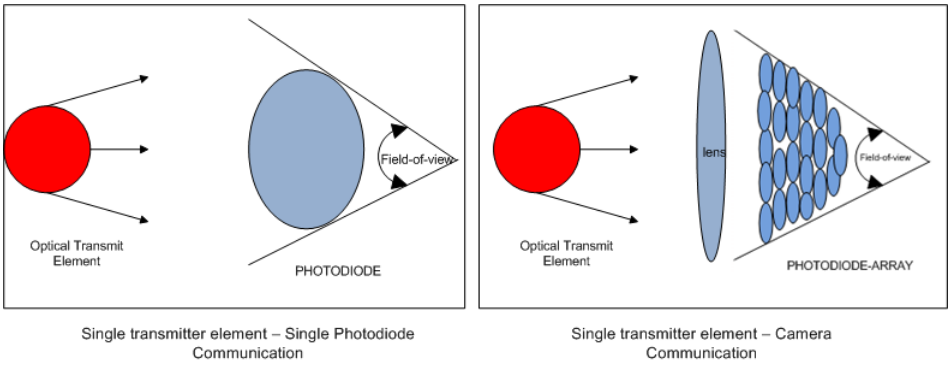


Figure 2.6: LED-Photodiode/Camera Communication Illustration

2.3.1 Array receiver v/s Single photodiode receiver

Photodiode arrays of a camera can provide a wide receiver field of view that allows for node mobility without the need to realign the receiver. Yet, by virtue of the camera design, each single photodiode element has a very narrow field of view, allowing high gain communication. The camera lens creates the effect of each photodiode being angled to a slightly different part of the scene, so that the combination of all diodes generates an image with a wide field of view. Other research groups have recently proposed variations of such designs [9]. For example, if larger receiver sizes are practical, the lens can be eliminated by using a photodiode array on a spherical receiver structure [16].

Apart from allowing more mobility array receiver also promises to achieve higher capacity than conventional optical wireless systems in a mobile setting where ranges greater than tens of meters are required. We justify this claim based on our capacity comparison between a photodiode array receiver and conventional optical receiver with only one photodiode element as shown in Fig. 2.6. We analyze a stationary communication model where a single LED with output power P_t transmits to an optical receiver over a wireless channel as shown in figure 2.2. This is a conservative model, because it does not include the effect of scene noise due to motion and achievable gains from multiple parallel transmission (from multiple LEDs). The two types of optical receivers we consider in our analysis are, (a) a conventional photodiode receiver and (b) a photodiode array (camera) receiver.

In an optical wireless channel, since the frequency of the optical signal is very large

compared to the rate of change of the impulse response, multipath fading and doppler shift are negligible. As described by Kahn and Barry [5], the received signal power follows $P_r = (RhP_t)^2$ where h is a channel parameter called channel DC gain and R is the receiver's responsitivity or the optical power to current conversion ratio. However, the received signal is corrupted by noise from the optical channel which is typically dominated by shot noise from background light sources and modeled as an additive white Gaussian process (AWGN) with a two sided power spectral density per unit area $S(f) = qRP_n$ [5, 17]. Here, q is the electron charge and P_n quantifies the power in background light per unit area. Hence, for a receiver sampling rate of W, the noise power is $P_N = qRP_nAW$ where A is the area of the photodiode. The signal to noise ratio for a single LED-single photodiode communication is,

$$SNR_{pd} = \frac{P_r}{P_N} = \frac{\kappa P_t^2 d^{-4}}{qRP_nAW} \tag{2.8}$$

where κ is a function of parameters such as the LED's lambertian radiation pattern, irradiance angle, field-of-view and optical concentration gain of the receiver [5].

Applying the model to the photodiode array receiver, we observe that the key difference between a conventional photodiode receiver and an array receiver lies in the detector area. When using the array, we assume the receiver can select the subset of diodes that actually observe a strong signal from the transmitter. This effectively reduces the detector area size and consequently reduces the noise. For the camera receiver (with a fixed-focus setting of the camera lens), we estimate the area of the array actually used through perspective projection [7]. Given a focal length f, a round LED of diameter l and the distance d between camera and LED, the LED will occupy a circle of diameter $l' = \frac{fl}{d}$ on the photodetector array. To conservatively account for the quantization effects, we assume that it will occupy a square area of size l'^2. This noise reduction gain is, however, limited by camera resolution. When the LED moves away from the camera, the projected diameter l' will eventually become smaller than the size of a photodiode. From this point on, the camera cannot further reduce the number of photodiodes that are used in the reception process and its performance becomes

similar to a single conventional photodetector (having the size of one pixel). We refer to distance where the LED generates an image that falls onto exactly one pixel as the critical distance $d_c = fl/s$, where s is the edge-length of a pixel.

Following this analysis, the signal to noise ratio for a single LED-photodiode array(camera) communication is,

$$SNR_{cam} = \begin{cases} \frac{\kappa P_t^2 d^{-2}}{qRP_n W f^2 l^2} & \text{if } d < d_c \\ \frac{\kappa P_t^2 d^{-4}}{qRP_n W s^2} & \text{if } d \geq d_c \end{cases} \tag{2.9}$$

We observe from equations (2.8) and (2.9), for $d < d_c$, that an array receiver has gain in SNR over a single photodiode receiver to the order of d^2. Thus at larger distances array receivers would be more resourceful than a single photodiode receiver. Also for $d > d_c$, though the array receiver is equivalent to a single photodiode in performance, in camera receivers the gain in performance can be achieved by reducing the pixel size s which is not possible with a single photodetector. For camera receivers since current off-the-shelf camera implementations are more limited in sampling rate (which equates to frame rate in camera) than photodetectors, a camera system will likely achieve even higher SNRs than a photodetector with a high-sampling rate. The lower framerate, however, also directly limits achievable rates.

To understand this tradeoff, given that the noise model is AWGN, we plot the Shannon capacity from 2.7, over a range of distances in figure 2.7 for a single photodiode receiver and three different camera receivers. Here, we use $W_s = 1$ (for camera we sum the total energy in the reception area) and neglect interference and consider only SNR instead of $SINR$. We set the sampling rate at 100MHz for the photodiode and 1000fps for the Basler Pilot piA640 machine vision camera & 100fps for SONY PS3eye webcam (two off-the-shelf cameras which use a CCD image sensor). We also consider a hypothetical camera which could sample at a rate of 1M fps. The parameter values underlying this result are summarized in Table 2.1. The graph shows that even at the low sampling rates of a webcam the camera system can still outperform the single photodiode due to its SNR advantage at larger distances. Moreover, the capacity of the

camera system can be increased considerably by using an array of LED transmitters (appropriately spaced) where the capacity at short distances can be scaled by a number equal to the number of LEDs and in some cases at longer distances too. We also see that the capacity of a camera system is more consistent over distance than for a single photodiode system for which it falls off rapidly (relatively) over distance.

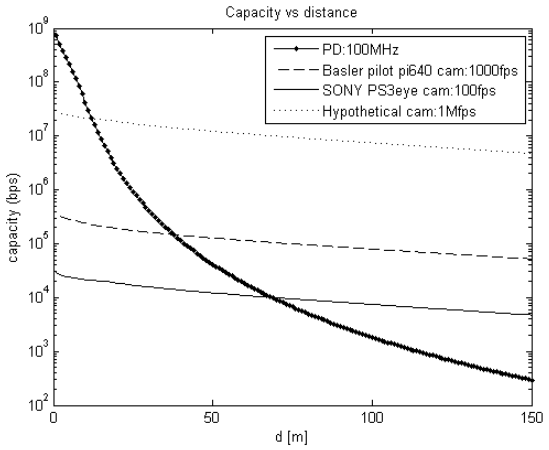


Figure 2.7: Capacity versus distance for Photodiode and Camera (visual MIMO) receivers

To further illustrate the array receiver advantage of eliminating noise by selecting only the photodiodes that receive the signal, we conduct an experiment with a blinking LED positioned 2m from the camera. The camera recorded a sequence of images in this completely stationary scenario. Figure 2.8 shows two histograms of the mean pixel value, one computed over a 10×10 area centered on the LED and one computed over the complete 640×480 image. These represent a single photodiode approach and a camera with the ability to eliminate background noise as discussed. The figure shows that in the first case the on and off state can be clearly distinguished through pixel values while in the second case the distinction is difficult since the signal is masked by shot noise.

Note that in a mobile transmitter-receiver scenario the camera's SNR gain (and hence the capacity gain) over a single photodiode can be expected to be pronounced because of scene noise, for example in a situation where the 'scene' has a strong reflector such as a white body. By extracting only those areas of the image that observe a strong transmitter signal, a camera can also selectively eliminate these distractors (noise) which is not possible with a single photodiode.

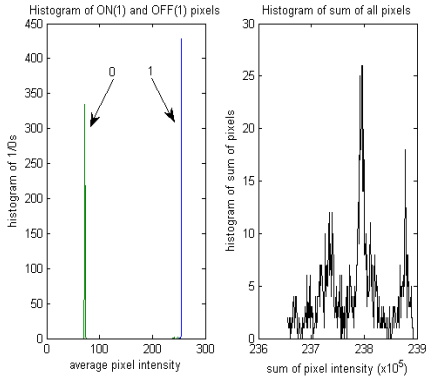


Figure 2.8: Histogram plots of Basler Pilot piA640 camera snapshots in medium sunlight(left:10×10, right:640×480)

Parameter	**PD**	**B**	**S**
P_t[mW]	100	100	100
$FOV\psi$[deg]	50	50	50
$A[mm^2]$	15.7	15.7	15.7
$P_n[mW/cm^2]$	600	600	600
l[mm]	6	6	6
f[mm]	–	21	6.5
$s[\mu]$	–	7.1	6

Table 2.1: Table of parameter values for photodiode and camera(**PD** Photodiode,**B** Basler Pilot piA640, **S** SONY PS3eye)

2.3.2 Multiplexing and diversity gains in visual MIMO

To quantify the perspective dependent multiplexing and diversity gains in visual MIMO we use the channel capacity of the visual MIMO channel as a metric which is given as,

$$C = \begin{cases} \sum_{k=1}^{K} W log_2(1 + SNR_{cam,k}) & \text{if } d \leq d^* \\ W log_2(1 + \sum_{k=1}^{K} SNR_{cam,k}) & \text{if } d > d^* \end{cases} \tag{2.10}$$

$$SNR_{cam,k} = \frac{\sum_{\forall I_k(i,j)=1} (h_k(i,j)x_k)^2}{\sum_{\forall I_k(i,j)=1} n_k^2(i,j)} \tag{2.11}$$

where W is the receiver sampling rate (camera frame-rate), d^* is the threshold multiplexing distance from equation (equation 2.6). $SNR_{cam,k}$ is the signal-to-noise ratio of the k^{th} LED at the camera receiver (2.9) which is expressed in terms of the

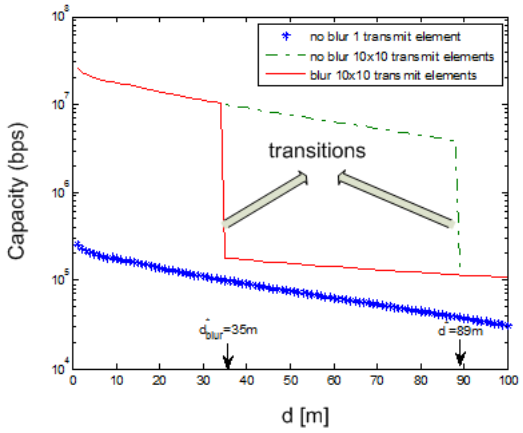


Figure 2.9: Visual MIMO channel Capacity versus distance ($\phi = 0$)

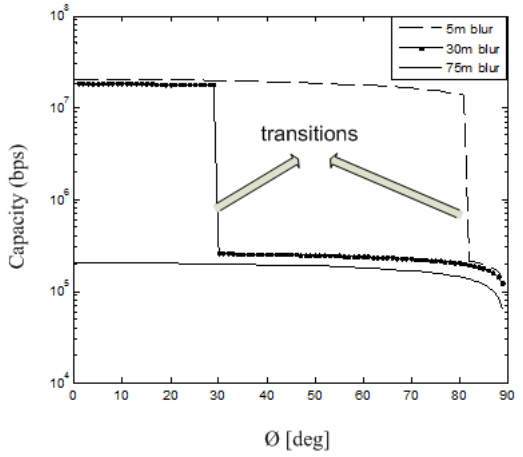


Figure 2.10: Visual MIMO channel Capacity versus angle (d constant)

transmit power x_k, the channel DC gain $h_k(i,j)$ from equation (equation 2.3) and AWGN noise $n_k(i,j)$ from equation (2.2). $I(.)$ is the indicator function, from equation (2.5).

We plot the channel capacity from equation (2.10), for an exemplary visual MIMO system, where the transmit elements of the LEA are light emitting diodes (LEDs) and the receiver is a machine vision camera (Basler Pilot piA640), over a range of distances d (Fig. 2.9) and over different viewing angles ϕ (Fig. 2.10). The underlying parameters used in our analysis are summarized in Table 2.1.

We can observe from the capacity plots in Figures 2.9 and 2.10 that a visual MIMO

system with no blur can achieve capacities of the order of Mbps even at long distances of about 90m. Blurring certainly reduces multiplexing range but still medium ranges of 30-40m are achievable at high data rates. The data rate gains at these distances are attributed to multiplexing where each LED sends an independent stream of bits over parallel channels. The transitions in the plot (for the multi LED cases) indicate the switch from multiplexing to diversity mode. The capacity gains due to diversity at the long distances, though may not be significant comparable to the multiplexing gains at shorter distances, are still close to an order of magnitude gain compared to the single LED system.

We also infer from our analysis that a visual MIMO system will have to switch between the multiplexing and diversity modes in discrete intervals based on distance and angle unlike RF MIMO where the gains in these modes could be achieved simultaneously but follow a continuous trade-off in performance. Moreover, a visual MIMO system will have to switch autonomously between these modes depending on the orientation of the receiver with respect to the transmitter in order to leverage the gains. This suggests that the throughput of visual MIMO links can be significantly improved through techniques that help the system adapt communication strategies with respect to receiver perspective.

The visual MIMO channel capacity is consistent over a wide range of viewing angles (small or large depends on distance). We see that the system can achieve large multiplexing gains at short distances and at almost all viewing angles which implies that the system would be robust to any misalignment between the transmitter and receiver. Its cleat that at large distances (of the order of 75m), due to the effect of lens blur, the LEDs may not be resolved easily even at $\phi = 0$ and hence at such distances where multiplexing will fail but using diversity over all angles can still offer an order of gain in data rates.

Such consistency in data rates over angular misalignment is important especially in mobile settings as the choice of multiplexing and/or diversity depends largely on the orientation of the mobile devices at each instance of time. This is in strong contrast to the RF systems (even MIMO) where the signal can drop significantly with mobility

especially when there is a deep fade in the channel or at high mobile velocities.

2.3.3 Visual MIMO versus RF

The key difference between a visual MIMO channel and radio wireless is that radio signals at the receiver are very random due to multipath fading effects. Multipath fading is a phenomenon where multiple copies – components created from reflections and scattering in the environment – of the same (transmitted) signal are received. Visual MIMO channel can be characterized, predominantly, as a deterministic channel as most of the signal energy lies on the LOS components while reflections and other multipath fading effects become negligible. This is because, light-waves undergo larger absorption in the wireless channel as compared to radio signals as they are much higher in frequencies. Therefore, the energy of the reflected light signal is almost negligible when it reaches the receiver. This means that the information capacity of radio systems is largely affected by fading and path-loss, while perspective (distance and angle) largely affects capacity of visual MIMO channel.

Radio systems today, in general, can achieve higher data capacities because of higher bandwidths available in today's radio systems compared to the available frame-rates in off-the-shelf cameras. For example, a Rayleigh fading radio channel can achieve close to 10s of Mbps [4] at a 10m distance (SNR close to 20dB) and bandwidth of 1MHz. A custom high-speed camera at 1 Million fps can achieve similar capacities at that distance, however, it may not be possible with off-the-shelf cameras as the frame-rates will few orders of magnitude lesser and the quantization limitations in recording the digital pixel intensity value will limit the maximum number of decodable bits per pixel. For example, an 8 bit RGB camera can achieve a maximum of 24 bits per frame, or 720bps with a 30fps commercial camera. Moreover, the additional processing required to eliminate noise from the images, address distortions due to motion, blur effects due to focus imperfections will eventually limit the throughput of camera communications. Even if we were to assume no distortions in the visual channel, and that it is only affected by background noise, then the capacity of a camera receiver would be slightly lesser than radio systems. This is because, the optical channel is also affected by shot

noise in additional to thermal noise, while radio channels do not have this effect.

In general, we learn that visual MIMO with custom cameras can achieve data rates comparable to radio systems with similar bandwidth availability. On the other hand, off-the-shelf cameras can help achieve nominal data rates, though much lower than radio. Visual channels can be used as side channels when there are outages in radio channels or where radio signals are highly regulated in use. In future, it may also be possible to strike a balance between the advantages of radio and visual communication to develop novel applications and system that use a hybrid of the two. One such technique will be discussed in Chapter 5 of this thesis.

2.4 Related Work

There is a large body of work in optical networking [18] and free space optics [19, 20] where the focus has largely been on stationary rather than mobile networks. Mobile free-space optics have also been designed for the military applications [21], however, they require mechanical trackers to enable mobility. Except for recent spherical FSO transceiver designs for mobile ad hoc networks [22] and optical satellite communications with physical steering [23, 24], mobile optical communications research has primarily focused on short range infrared communications for mobile devices [5, 9]. While earlier work has used cameras to assist in steering of FSO transceivers [25],the visual MIMO approach differs by directly using cameras as receiver that can facilitate to design an adaptive visual MIMO system that uses multiplexing at short distances but still can achieve ranges of hundreds of meters in a diversity mode. It exploits advances in CMOS imagers that allow higher frame rates compared to earlier CCD designs.

There is an exploding interest in using the visible light spectrum for communication [6, 26–30]. Low-speed audio communication systems using LED transmitters have been demonstrated [31]. In Japan, a consortium of 21 research groups called the Visible Light Communication Consortium (VLCC) has been formed to research into areas of VLC [28]. Since 2008, the Smart Lighting research group at Boston University [32] has been investigating visible light communication systems for indoor lighting and outdoor

vehicle to vehicle applications [33, 34]. The work so far generally uses photodiodes or custom image sensor circuity at the receiver to convert the optical signals to electrical signals. Though photo diodes can convert pulses at very high rates, they suffer from large interference and background light noise. This results in very low signal-to-noise ratios (SNR), which leads to the short range of typical IR communication systems, even with more sophisticated receiver processing and modulation techniques as studied in [35].

Saito et al. investigate the use of image-sensor receivers for inter-vehicle communications [17] using LED light emitters, and traffic light to vehicle communications has been investigated in [36]. Other work has investigated channel modeling [6] and multiplexing [37]. More recently, researchers of the MIT Bokode project [38] have applied computational photography to camera based communications. For shorter range systems [39, 40] show a MIMO approach for indoor optical wireless communication, [41] studied the capacity of a optical MIMO system and [42] details some work on space-time codes for optical MIMO. Earlier work by Kahn [9] investigates the use of multibeam transmitters and imaging receivers in infrared systems, very similar to MIMO in concept.

2.5 Conclusion

In this chapter we introduced the idea of visual MIMO where camera acts as receivers for information transmitted from light emitting arrays, and developed a channel model for visual MIMO. Visual MIMO allows communication range of hundreds of meters with a relatively wide field-of-view compared to free-space optics, thereby enabling a higher a degree of node mobility. Our analysis showed that even visual MIMO system using a toy webcam can achieve close to order of magnitude gains in bit-rate over a conventional photodetector receiver with the same field-of-view and there is significant room for improvement through more specialized image sensors. We observed that the gains in visual MIMO are highly perspective dependent, where multiplexing gains can be obtained at short distances while ranges of hundreds of meters can be achieved in a diversity mode. Our analytical results report - even in the presence of signal distortion

due to lens blur - channel capacities of the order of Mbps at short distances and of the order of hundreds of Kbps at medium to longer ranges for an exemplary visual MIMO system with 100 LEDs in an array. We also showed similar channel capacities for the same system over wide camera view angles. These results validate the premise that the MIMO gains in an optical MIMO system such as visual MIMO is primarily dependent on receiver perspective with respect to the transmitter in contrast to the multipath fading dependent gains in RF MIMO. We inferred that a visual MIMO system will have to switch between its multiplexing and diversity mode unlike RF MIMO where they can be achieved simultaneously but follow a tradeoff in performance. The consistency in data rates over a wide range of camera viewing angles is a positive indication that visual MIMO can enable mobility in optical wireless communication. In this chapter we emphasized that regardless of any type of modulation and transmission scheme, visual MIMO can still achieve significantly high data rates by leveraging some of the unique characteristics of the visual channel. Visual MIMO system still presents a broad spectrum of opportunities and challenges for mobile computing and networking research.

Chapter 3

Capacity of Screen-Camera Communication: A Visual MIMO application case-study

This chapter studies the fundamental capacity limits of an example use-case visual MIMO application where display screens are used as transmitters, which we will formally refer to as *screen-camera* communication. Communicating from screens to cameras could be particularly attractive in pervasive camera based applications, where such camera communications can reuse the existing camera hardware and also leverage from the large pixel array structure of display screens for high data-rate communication. In this chapter, we discuss our model of screen-camera communication that builds over the visual MIMO channel model chapter 1. However, the model presented in this chapter accounts for the reality in off-the-shelf cameras such as quantization limitations and frame-rate limitations. We use the model to predict the capacity of screen-camera communication and study the effect of receiver perspective (distance and angle to the transmitter) on capacity. We also calibrate and validate this model through lab experiments.

3.1 Camera Communications

Today, cameras are frequently used to read QR-codes, which can be considered as a form of visual communication wherein the camera acts as a receiver. The ubiquitous use of QR codes motivates building novel camera communication applications, where pervasive display screens could be modulated to send time-varying QR codes to be decoded by video cameras. The large pixel array elements of the screen and camera can be leveraged to send high volume of data through short time-varying 2D barcodes.

For example, a user could point a camera to a desktop PC or a smartphone screen displaying the time-varying code to download a file.

A camera channel is analogous to a RF MIMO channel where each pixel element of the camera acts as a receiving antenna and the light emitting elements as the transmit antennas. In RF MIMO, the signal quality at each receive antenna element is a function of the path-loss in the channel, multipath fading, and the interference from other transmit antennas — also called co-channel interference [4]. A camera channel has negligible multipath fading, but experiences path-loss in light energy, and interference (of light energy) from other light emitting elements, which manifest as visual distortions on the output of a camera; that is, the image. These distortions are primarily a derivative of the camera imaging process and can be modeled (deterministically) using classical camera imaging theory.

The signal quality at the camera receiver is also influenced by noise in the channel. Noise in camera systems manifests as spurious electric signal in the form of *current* on each camera pixel. Noise current is generated due to the photons from, environment lighting (includes ambient lighting) and from the transmitter and receiver imaging circuitry [43]. Noise current in a pixel is usually considered signal independent when the ambient lighting is sufficiently high compared to the transmit signal; for example, in office rooms or outdoors [9]. At the output of a camera, the noise current in each camera pixel is a quantized quantity and manifests as fluctuations in the intensity (digital value of the sensor output) of that pixel; the noise energy accumulated in each pixel can be quantified using the mean value of variance in the pixel intensity. As in prior works that modeled optical channels [6,9], in this work, we consider that the noise in a camera pixel is primarily from the background, and follows a AWGN characteristic (quantified through the AWGN noise-variance σ_n^2), and is uniform over the image sensor (photoreceptor).

3.2 Screen - Camera Channel

In screen-camera communication, information is modulated in the light intensity of the pixel elements of a screen transmitter that are received and decoded from the camera image pixel intensity at the receiver. The pixel intensity in a camera image is a digital quantity (most integrated cameras have 8 bit monochromatic depth (on each colour channel) where the values span 0 (dark)-to-255 (bright)) that is proportional to the amount of photon current generated on the pixel from the light energy accumulated over its area (the smaller the pixel area the lesser light intensity it accumulates). When the light emitting screen pixel is at the focus of the camera lens all the light rays from the screen pixel are focused onto a camera pixel and thus incurring no loss of energy on the pixel. When the screen pixel is perturbed (in position and/or orientation) from the focus of the camera or incurs path-loss in energy due to the finite aperture size of the camera lens, not all light rays converge on the camera pixel resulting in reduced accumulated energy and hence a lower pixel intensity value. The loss in the received light intensity on a camera pixel results in the visual deformation in size or shape of the imaged screen pixel; an effect that is termed as perspective distortion.

Loss in signal energy on a pixel is also attributed to the noise in that pixel. Noise in a camera pixel is primarily due to spurious photons (that do not belong to the transmitter) from the environment, which can be modeled as signal independent and AWGN. Noise from the transmitter and the camera imaging circuit are dependent on the generated signal (and that is transmitted), and thus depend on the transmitter and receiver specifications. However, unlike environment noise, this signal dependent noise can be estimated using one-time calibration mechanisms; camera noise modeling has been well studied in computer vision and solid-state electronics (CMOS) design literature. We reserve the discussions on effect of signal dependent noise on throughput of camera communications for future work.

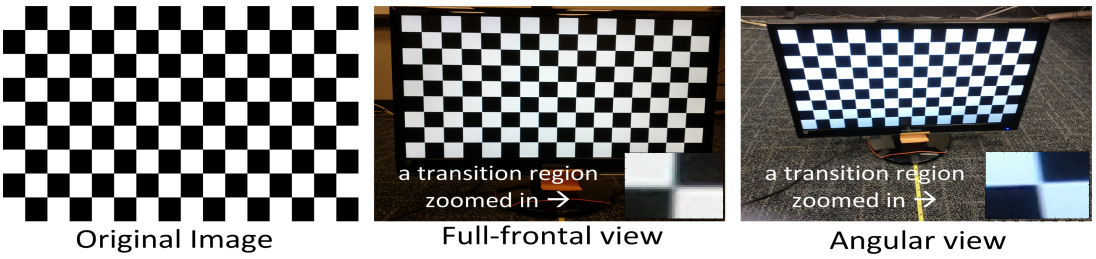


Figure 3.1: Illustration of perspective distortion in screen-camera channel. Imaged screen pixels are blurry, and reduced in size in full-frontal view and also in shape in angular view.

3.2.1 Perspective Distortions

Distortions that depend on the perspective of the camera are caused due to the nature of the camera imaging mechanism and manifest as deformation in size and shape of the captured object (the light emitting screen pixel) on the image, resulting in visual compression or magnification of the object's projection on the image. When the screen is at an out-of-focus distance from the camera lens (or at an oblique angle), these distortions become prominent and lead to interference between adjacent screen pixels on the camera image, what we term as inter-pixel interference or IPI. The combined effect of background noise and IPI degrades the received signal quality and hence reduces information capacity in camera channels.

For example, let us consider that blocks of pixels on a screen are illuminated by a chessboard pattern and imaged by a camera as shown in Fig. 3.1. We can observe that perspective distortions cause the screen pixels to deform in size when the screen is not at the focus of the camera, and in shape when it is not frontally aligned (viewed at an angle) with the camera.

Perspective scaling. If the screen pixel was at the focus, and assuming the screen and camera have the same resolution, it's image on the camera should occupy the same area as one pixel. But in reality, the light rays from the screen pixel may not end exactly on camera pixel boundaries and there is some area surrounding it that accumulates interference. This area of misalignment and the geometry of the imaged screen pixel will be perspective (distance and orientation) dependent and accounts for distortion due to perspective scaling of the pixel area.

Lens-blur. We can also observe from Figure 3.1 that the imaged screen pixels

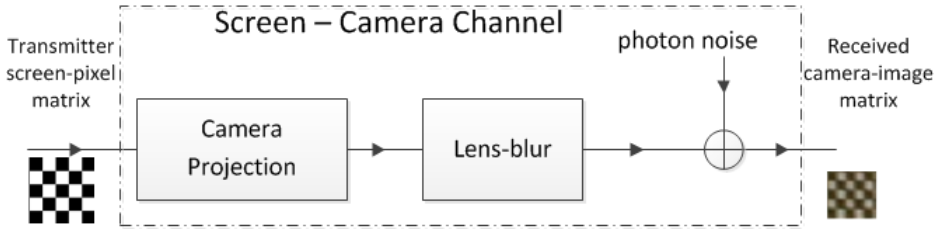


Figure 3.2: Screen - Camera Channel Model

are blurry, especially at the transition regions between white and black blocks. This blur effect is attributed to the camera lens and more formally termed as lens-blur. This blur effect is typically modeled in camera imaging theory using the point-spread function (PSF) [**?**], which represents the response of an imaging system to a point source. In the screen-camera channel this translates to distorting the pixels at the transition regions between brighter (high intensity) and darker (low intensity) pixels, and leads to interference (IPI) between neighboring pixels, as seen in Figure 3.1. Since the area and the maximum energy that can be sampled in each camera pixel is finite, IPI leads to an effective reduction in signal energy per pixel.

3.3 Modeling Perspective Distortion Factor

In this work, we model the perspective distortions in the screen-camera channel as a composite effect of signal energy reduction due to perspective scaling of pixel area owing to camera projection, signal energy reduction due to lens-blur, and background photon noise, as shown in Fig. 3.2. In this regard, we consider that the signal energy on each pixel is weighted by an average perspective distortion factor α, that represents the effective area scaling (down) due to perspective and lens-blur in the camera imaging process, while the rest of the light-energy on the pixel is from ambient photon noise. We define this factor such that it takes values in $0 \leq \alpha \leq 1$, where $\alpha = 1$ indicates that the screen pixel is at the focus of the camera and also incurs no signal reduction due to lens-blur, and $\alpha = 0$ indicates that no part of the screen-pixel gets imaged on the camera pixel.

3.3.1 Perspective scaling

Let α_p represent the perspective scaling of the area of an imaged screen pixel when perturbed from camera focus. We model this perspective scaling factor using camera projection matrix [44] which maps the location of the screen pixels from the world coordinate frame to the camera coordinate system. We have discussed the derivation for a general expression for α_p using camera projection theory in section 3.9. In the simplest case, where the screen and camera are perfectly aligned at distance d, this factor can be expressed as,

$$\alpha_p = (\frac{f_{cam} s_t}{s_{cam} d})^2 \tag{3.1}$$

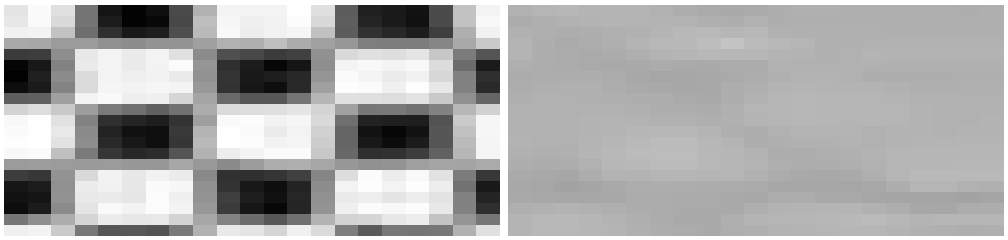
where f_{cam}, s_t are the focal length of the camera and side-length of the screen pixel, respectively. We can observe from equation (3.1) that, $\alpha_p = 1$ when the camera is at the focus ($d = f_{cam}$) and if $s_{cam} = s_t$. However, in reality, the physical size of a screen and camera pixel may not be the same. In our system, we assume that the focal point is at a distance $d_f = \frac{f_{cam} s_t}{s_{cam}}$ to the screen; which we term as *focal-distance*.

3.3.2 Lens-Blur

As discussed earlier, lens-blur causes the signal energy to leak outside the area of a single pixel. Camera lens-blur, characterized by the PSF, can be approximately modeled as a 2D gaussian function [**?**, **?**], where the amount of spread in area is quantified using its variance σ^2_{blur} (a large variance indicates more blur[1]). In our model we account for lens-blur distortion using the factor $\alpha_b = (2\sigma_{blur})^2$, to account for the spread in area over two dimensions of the square pixel. If s_{cam} is the side length of a camera pixel, then the effective signal energy on that pixel will be proportional to $s^2_{cam} \frac{1}{1+\alpha_b}$. We treat this signal energy reduction is proportional to this reduced pixel area over which the signal accumulates.

In this regard, we consider α, an average distortion in each pixel of the camera image

[1]For an ideal pin-hole camera energy spread over a pixel would be uniform and hence σ^2_{blur} is infinitesimally small



(a) handheld (b) motion

Figure 3.3: Illustration of motion blur on images of a screen displaying a chessboard pattern, taken by a hand-held camera (a) and when camera is in motion (b)

to quantify perspective distortion. We express α as the effective pixel area reduction due to perspective scaling factor α_p on the reduced pixel area due to lens-distortion $\alpha_b = 4\sigma_{blur}^2$, as

$$\alpha = \alpha_p \times \frac{1}{(1 + \alpha_b)} \tag{3.2}$$

In reality, the physical size of the screen pixel may not exactly be matched with that of the camera image sensor. This can cause an imaged screen pixel not to align with a camera pixel, even if the screen pixel were at the camera focus. Such misalignments will cause a deviation in the distortion factor for each pixel as the perspective changes. However, such deviations can be assumed to be negligible when considering an average distortion factor over the camera image.

3.3.3 Motion Blur

Screen-camera communication applications would typically involve some degree of motion, for example, when the camera is hand-held, or when the camera or screen is on a moving vehicle. Motion due to hand-shakes or lateral movements can cause dynamic change in perspective between the screen and the camera. In such cases, one can assume some vibrations on the pixels, especially when the camera is not stable, where the pixels seem to interfere with each other, eventually causing a blurry visual effect on the image; formally known as motion-blur in computer vision [**?**].

Motion-blur primarily arises due to movement within or between camera frames. Smartphone cameras are usually hand-held and vibrations caused due to hand motion

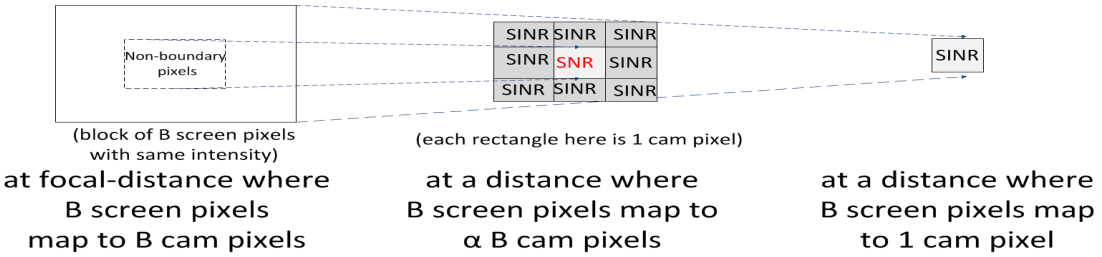


Figure 3.4: Illustration of interference between pixel-blocks due to perspective distortion for SINR computation

can cause motion blur but are usually much less than those when the screen or camera is in motion. Cameras equipped in vehicles may suffer from more blur compared to hand-held scenarios as camera sampling may be too slow when compared the speed of motion. Figure 3.3 shows an example of camera snapshots of a screen imaged when camera is (a) hand-held, and (b) in motion. We can observe from these snapshots the distortions due to motion blur leading to inter-pixel interference.

Cameras today are equipped with very effective *motion compensation* capability which compensate motion blur through a filtering mechanism called *de-blurring*. De-blurring [**?**] is a technique that is commonly used to mitigate the effect of blur on the image by applying a filter that inverts the effect of blur on the image. The quality of the de-blurred image will largely depend on the effectiveness of the de-blurring filter as well as the amount of induced motion/vibration on the pixels. Imperfections in the de-blurring process can also lead to signal quality reduction compared to an ideal (static screen and camera) scenario. If the motion is fast then the camera may not be able to expose to the entire screen pixel and hence causing the signal energy to spread over many pixels and result in a more blurry image as shown in Figure 3.3 (b). We note that the α_b factor in the perspective distortion factor α quantifies the effective blur in a camera pixel.

3.4 Signal-to-Interference Noise Ratio in Screen-Camera Channel

We quantify the quality of the signal at the camera receiver in the screen-camera channel using the average SINR per pixel,

$$SINR_{\alpha} = \frac{\alpha P_{avg}^2}{(1-\alpha)P_{avg}^2 + \sigma_n^2} \tag{3.3}$$

where, P_{avg} denotes the average transmit pixel intensity. For example, a screen-camera system using black (digital value 0) and white (digital value 255) intensities for transmission will have $P_{avg} = 127.5$. By using the digital value of the average signal P_{avg}, instead of its analog equivalent (pixel photon-current squared), our model accounts for the quantization limitations in cameras. The $1 - \alpha$ term in equation (3.3) quantifies the fraction of the pixel area affected by interference.

3.4.1 Pixel blocks

A small value of α indicates that more screen pixels interfere on one camera pixel. In reality, screen pixels are very closely spaced (fraction of a mm), and so, IPI will be inevitable even at short distances resulting in low SINRs. A potential solution is to leverage the MIMO structure of the screen-camera channel, by grouping multiple screen pixels in a block, such as a 2D barcode, to transmit with same intensity, and combine such pixels from the camera image to improve SINR. This technique, in principle, is similar to diversity combining used in RF MIMO. Pixel-blocks merely represent that a group of antennas are used to transmit the same intensity, to improve the SINR at the receiver. By using pixel blocks, we draw analogies of the screen-camera channel to an equivalent MIMO system. This is different from considering multiple-level modulation or coding to improve communication throughput. In this paper we are primarily interested in determining the bounds on the information capacity which by definition is independent of the type of modulation or coding used.

Pixel blocks are effective in reducing the impact of misalignments, and lens-blur, as these effects become smaller as one block covers more pixels on the camera and only affect pixels near the boundary as shown in Fig. 3.4. The SINR can be enhanced by considering averaging the signal energy over such blocks of pixels.

As a convention in our model, we treat a pixel block as a boundary block if it is not all surrounded by blocks with same intensity. Such a structure minimizes the

'interference'for a non-boundary pixel, and is negligible when the camera and screen are static with respect to each other. In this case, even for non-zero blur or pixel misalignment, since the same signal adds-up on the pixel, it enhances signal energy of that pixel; in which case the SINR of that pixel converges to the average-SNR.

In general, the expression for the average SINR per imaged block in a screen-camera channel, using B pixel square blocks of a screen can be given as,

$$\begin{aligned} SINR_{blk}(\alpha, B) &= \gamma_1 SINR_\alpha + \gamma_2 SNR_\alpha \forall \alpha B > 4 \\ &= SINR_\alpha \qquad\qquad\qquad \forall \alpha B \leq 4 \end{aligned} \tag{3.4}$$

where $SINR_\alpha$ is from equation 3.3, $SNR_\alpha = \frac{\alpha P_{avg}}{\sigma_n^2}$, and the coefficients $\gamma_1 = 4(\sqrt{\alpha B} - 1)$ and $\gamma_2 = (\sqrt{\alpha B} - 2)^2$ represent the number of boundary-blocks and non-boundary blocks, respectively. Here, $\min B = 4$ (i.e. 2×2 pixels), and $\alpha B \leq 4$ indicates that each B pixel block projects onto a maximum of 1 camera pixel area while $\alpha B > 4$ indicates that the block projects onto multiple camera pixels.

3.5 Capacity Under Perspective Distortions

Recalling the capacity expression from equation (2.7), we can express the capacity of screen-camera communication in bits/sec as,

$$C_{cam}(\alpha) = \frac{W_{fps}}{2} \alpha ||R_{cam}|| log_2(1 + SINR_\alpha) \tag{3.5}$$

where $SINR_\alpha$ is the signal-to-interference noise ratio from equation (3.3), $||R_{cam}||$ denotes resolution of the camera and W_{fps} denotes the frame-rate of the camera in frames-per-second. The camera frame-rate, and hence bandwidth, is halved (following Nyquist sampling theory) to avoid the mixed frames caused by aliasing resulting from the synchronization mismatch between screen updates and the camera sampling. The term $\alpha ||R_{cam}||$ represents the total number of camera pixels that contain the image of the screen pixels, and directly corresponds to the spatial-bandwidth term W_s in equation 2.7. This is very different from RF MIMO, where, all the receiver antennas can potentially receive the signal, independent of distance between the transmitter and

receiver. In a camera receiver, due to its LOS nature, the signal from each transmit element is always limited to a finite number of, but never all, receive elements.

3.5.1 MIMO throughput

The capacity in equation 3.5 represents the upper bound on the total number of bits that can be communicated with negligible error from one screen pixel to a camera pixel. Grouping pixels into blocks improves the SINR and reduces bit errors, but the effective data throughput scales down as the number of parallel channels are reduced. This behaviour is similar to the classical multiplexing-diversity tradeoff in RF-MIMO [**?**]. If $T_{blk}(\alpha, B)$ represents the MIMO capacity or maximum throughput of screen-camera communication for block-size B, at distortion factor α, then

$$T_{blk}(\alpha, B) = \frac{W_{fps}}{k}(\frac{\alpha||R_{cam}||}{B})log_2(1 + SINR_{blk}(\alpha, B)), \tag{3.6}$$

where $\frac{\alpha||R_{cam}||}{B}$ represents the number of parallel channels for multiplexing, and $SINR_{blk}(\alpha, B)$ is from equation (3.4). In practice, to minimize detection and decoding errors, the camera frame-rate has to be synchronized with the modulation rate of pixel intensities on the screen as well as the refresh rate of the screen (typically 120Hz). The factor k in equation 3.6 corresponds to the oversampling factor to address the asynchronism between the screen (data) update rate and the camera sampling rate. It implies that a minimum of k temporal samples of the camera pixel are required for reliable decoding. Synchronization of cameras for communication is challenging due to the jittery nature (owing to software limitations and hardware design errors) of the frame-sampling using CMOS sensors that are widely used in mobile devices today.

3.6 Experimental Calibration and Validation

In this section we describe the experiments we conducted to validate our screen-camera channel model. The key motive of these experiments was to determine the channel capacity in a real screen-camera channel.

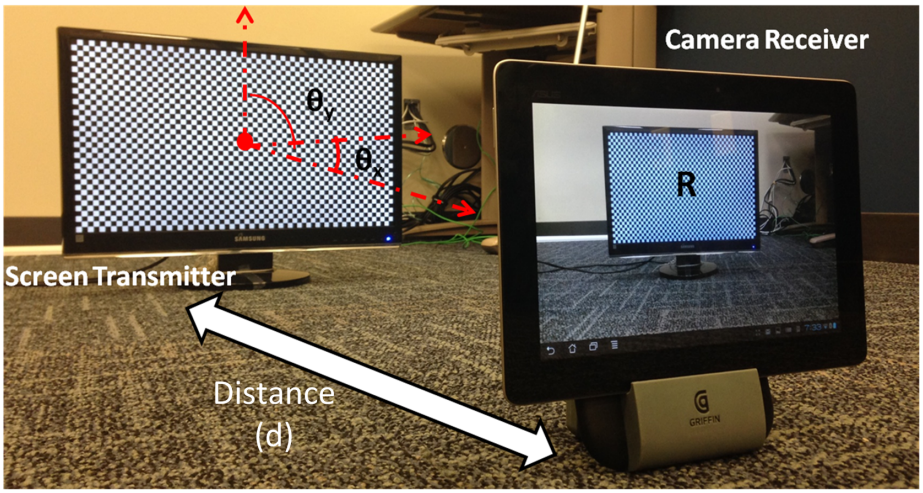
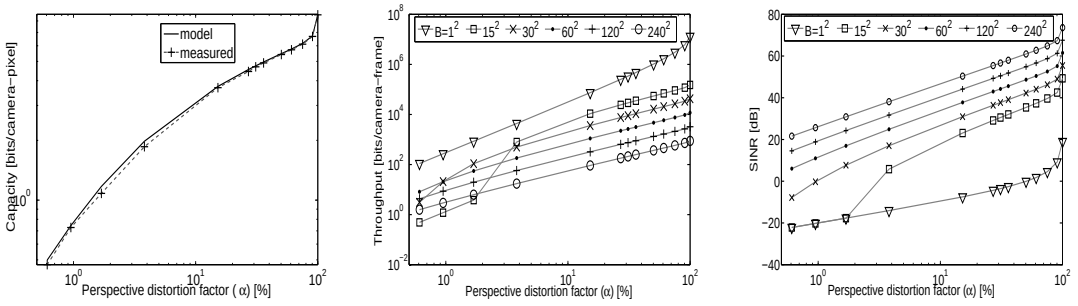


Figure 3.5: Experiment setup showing LCD screen displaying black and white blocks of $B = 60 \times 60$ pixels each

Measured channel capacity. It is a fact that it not possible to measure capacity of any communication channel directly, hence we aim to determine capacity indirectly by substituting the measured SINR, perspective distortion factor α and noise power into the analytical capacity expression derived in (3.5). Our experiments were aimed at measuring these specific parameters that aid in determining capacity values for an example test channel that we considered. However, we note that these experiments as well as the findings can be applied to a generic camera communications channel – with appropriate specifications of the transmitter and receiver considered. In this work, we estimate capacity of screen-camera channel by substituting the measured values of $SINR_\alpha$, perspective distortion factor α, and noise variance σ_n^2 in equation 3.5. The measurement procedure for α, $SINR_\alpha$ are explained in detail in sections 3.6.4, 3.6.5 respectively.

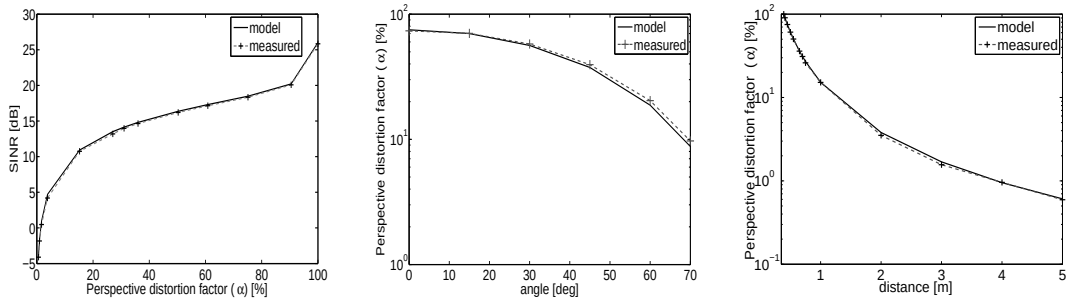
3.6.1 General Experiment Methodology

The experiment setup, as shown in Fig. 3.5, consisted of a 21.5inch Samsung LCD screen monitor of resolution $R_s = 1920 \times 1080$ pixels, that served as the screen-transmitter, and a 8MP camera of a ASUS Transformer Prime tablet (that ran Android OS version 4.1), that served as the camera receiver. The camera was operated at a resolution of



(a) (b) (c)

Figure 3.6: (a) Capacity in bits/camera pixel ($C_{campixel}(\alpha)$) for different perspective scaling (α) of screen image on camera (b) Throughput in bits/frame v/s α for different blocksizes (1 frame = R_{cam} pixels, $B = 15^2$ means 15×15 pixel block on screen) (c) SINR per block v/s α for different blocksizes B



(a) (b) (c)

Figure 3.7: (a) SINR for different perspective scaling (α) of screen image on camera (b) Perspective distortion α v/s angle between screen and camera (c) Perspective distortion factor α v/s distance between screen and camera

$R_{cam} = 1920 \times 1080$ and with no image compression. Exposure setting and white-balancing on the camera were set to auto (default setting in Android devices). All the experiments were conducted under the same environment lighting conditions with the measurements taken indoors in a lab-conference room setting equipped with fluorescent ceiling lighting. We fixed the screen and tablet onto stands so as to ensure the least amount of error in the measurement of distance and angle between the tablet and camera image planes. The raw dataset for our analysis consisted of image snapshots of the screen, displaying a chessboard pattern (blocks of B pixels each), captured by the tablet's camera at resolution of R_{cam} pixels using a standard Android image capture application. The camera parameters were obtained through a well known calibration

toolbox [45]. The pixel-intensity of a white block was set to 255 and the black at 25[2] on the screen (the average intensity $P_{avg} = 140$). The image datasets consisted of 100 snapshots of the screen displaying the chessboard pattern, with the ceiling lights ON (an another dataset with lights OFF), at a set of distances, angles, and block-sizes. We changed angle between screen and camera by rotating the screen with respect to the X axis; distortions can be considered symmetrical on X and Y axis.

Table 3.2 summarizes the list of measured parameters from our experiments, along with the screen and camera specifications.

3.6.2 Channel Capacity

We evaluate capacity in bits per camera pixel as $C_{campixel}(\alpha) = \frac{C_{cam}(\alpha)}{\frac{W_s}{2}||R_{cam}||}$.

Capacity v/s Perspective distortion factor

We plot the measured capacity in bits/camera-pixels for different perspective distortion factor values in Figure 3.6 (a) along with the analytical values, and observe a good fit (maximum error margin of 3%) between the two. The distortion factor α on the x-axis is comprehensive of composite distortion due to perspective scaling as well as blur. We can observe that, about 1bit/camera pixel is achievable even when the screen is perspectively scaled onto only 15% on each dimension ($\alpha = 2\%$) of the camera image. For the LCD screen-tablet camera system we used, this translates to a distance of 2.6m. At a sampling rate of 30fps[3] and at a resolution of 1920×1080, a data-rate of 31Mbps is achievable from an average-sized LCD monitor and a tablet camera. Assuming all parameters are the same, except the size of the screen is doubled, the same data-rate can be achieved at twice the range. Such data-rates are even sufficient for streaming a video.

---

[2]Due to the screen's residual back-lighting, intensities in [0,25] range did not cause any change in screen brightness

[3]Typical frame-rate on smartphone/tablet cameras is 30fps. IPhone 5S has a 120fps capability [**?**]

Throughput with Block-size

We plot the screen-camera communication throughput from equation 3.6 in bits-per-frame ($\frac{T_{blk}(\alpha,B)}{kW_{fps}}$) for different values of perspective distortion factors, and block sizes B, in Figure 3.6 (b). We can observe from Fig. 3.6 (b) that capacity falls of steeply as α becomes smaller for smaller block-sizes; for example, at $B = 15^2$ and 30^2. The trend can be attributed to the low SINR at those perspectives as IPI increases due to the dense arrangement of bits (pixels carrying unique information). A block-size of 1 does not follow this trend as the gain from the capacity scaling due to more number of parallel channels compensates for most of the loss in SINR, however, trading-off with receiver complexity to detect the very low SINR signal.

Throughput comparison with existing prototypes

We compare our MIMO capacity estimates ($T_{blk}(\alpha, B)$) with the throughput of existing prototypes of screen-camera communication. In PixNet [46], bits are modulated onto LCD screen pixels that are decoded by an off-the shelf point and shoot camera. PixNet uses OFDM for modulation and adds (255,243) reed-solomon coding for error correction. Consistent with the definition of a block in our model, PixNet uses a block-size of 84×84. PixNet was evaluated using a 30inch LCD screen as the transmitter and 6MP CCD camera at the receiver, and up-to a maximum distance of 14m. The authors also reported the throughput from their implementation of QR codes, which we will call QR-P. The QR-P uses a version 5 QR code with a block size of 5×5 pixels, and that encodes 864 bits per QR code. On the other hand, COBRA [47] uses color barcodes to communicate between smartphone screen and camera, and was evaluated up-to a maximum distance of 22cm, and with a blocksize of 6×6 pixels. The authors of [47] have also implemented a smartphone (receiver) version of PixNet, which we will call PixNet-C, where the settings remained the same as original PixNet system.

COBRA	PixNet-C	PixNet	QR-P
4.5x	3x	2.5x	7x

Table 3.1: Ratio of capacity over existing prototype's throughput (3x indicates the existing prototype is 1/3rd of capacity)

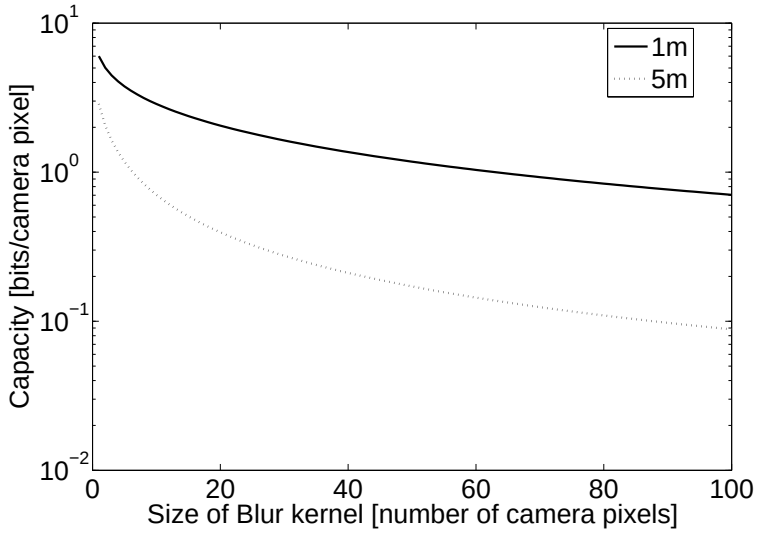
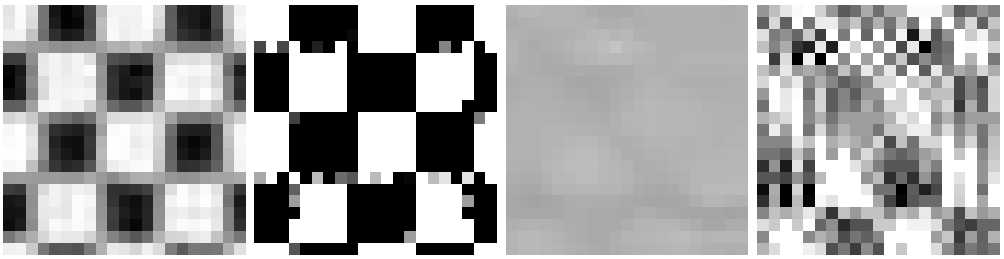


Figure 3.8: Capacity v/s blur

In table 3.1, we report the ratio of throughput from equation 3.6 to the throughput of the these prototypes, for the same parameter settings, of blocksize and α as in their existing implementations. Our estimates indicate that there is room for atleast 2.5x improvement in throughput when compared to capacity. The discrepancy in throughput in these existing prototypes can be attributed to different parameter choices. For example, PixNet uses OFDM modulation and coding which add communication overheads, which have to be incorporated in a limited spatial bandwidth available on the screen. COBRA also incurs loss in throughput due to coding overheads, and additionally the small block size allows for more interference, reducing SINR. COBRA minimizes blur by using repetitive colour patterns and intelligent placement of those patterns on the screen. While this strategy minimizes the effect of interference from neighboring pixels, the repetition causes under-utilization of the spatial bandwidth. In general, our findings, supported by these exemplar comparisons, open up interesting questions in the design space for improving information throughputs of screen-camera communication systems.



(a) handheld (b) deblurred hand-held (c) motion (d) deblurred motion

Figure 3.9: Illustration of motion blur and deblurring on images taken by a hand-held camera (a) and (b), and when camera is in motion (c) and (d)

3.6.3 Motion-blur experiments

To understand the effect of blur alone on the capacity we first plot the measured capacity $C_{campixel}(\alpha)$ at a fixed perspectives (distance of 1m where α =0.5 and 5m where α =0.5 and at angle=0) in Figure 3.8. We observe that blur can significantly affect capacity, for example we can observe that the capacity drops drastically when the blur levels are high even when the perspective scaling is only 50%. We observe that the capacity drop is steeper at long distance. We note that a blur kernel of size 1 pixel indicates no blur and at this perspective (α value of 1) the capacity is 6bits/camera-pixel for the distance and angle between the screen and camera in this experiment.

To understand the effect of motion blur on the signal quality and the effectiveness of de-blurring, we conducted an experiment where we captured a video stream of the screen displaying a chessboard pattern with white and black blocks of size 15×15 pixels. During the course of this experiment the camera was hand-held for one case, and the other case, the hand-held camera was intentionally moved (in a horizontal waving pattern) at a nominal speed approximately equivalent to when the user is walking. The distance between the screen and camera was 1m; at this distance only 50% of camera image is occupied by the screen transmitter pixels. We then applied a Weiner filter based deblurring function available in MATLAB [**?**] to each of the 100 consecutive images from the video-streams in both cases (see Figures 3.9 a-d). Similar to the previous experiments, we then estimated the capacity of screen-camera communication for these two cases by estimating the average perspective distortion factor and the

average SINR from the deblurred images. Our estimates indicated a capacity value of 5bits/camera-pixel for the hand-held case and about 2bits/camera pixel for the motion case. With reference to Figure 3.8 our findings from this experiment indicates that even when the camera is hand-held the capacity of screen-camera communication can be reached as close to as it is when the camera is stationary – with the motion de-blurring features available in off the shelf cameras. For the motion case, without de-blurring, the capacity is almost zero due to the large number of bit errors due high inter pixel interference. However, we observe that de-blurring can help achieve a reasonable data capacity. From this experiment we infer that by using a simple filtering operation the capacity can be improved to a reasonable amount. We also infer, based on Figure 3.8, that the amount of blur is approximately 1 pixel for the hand-held case and about 15 pixels for the motion case, in each dimension.

3.6.4 Perspective Distortion Factor

The objective of this experiment was to determine the perspective distortion factor α from our measurements to estimate capacity. Since α quantifies the relative area occupancy of the screen in the camera image, we measured the average distortion factor as,

$$\alpha_m = \frac{||R||}{||R_{cam}||} \frac{1}{(1 + 4\sigma^2_{blur})} \tag{3.7}$$

where $||R||$ represents to the total number of camera pixels that correspond to the imaged screen pixels, and R_{cam} is the resolution of the camera. In figures 3.7(b) and 3.7(c) we plot α_m as a function of angle and distance, respectively. As can be seen from these plots the measured spatial-bandwidth fits well with the model (maximum error margin of 1.5%). The α_m reported here is the perspective distortion factor for our LCD - tablet (camera) channel. The distance and angle at which $\alpha_m = 0$ in these plots can be construed as the communication range of a system with the same screen and camera parameters. For example, for a screen with 10x the size (a billboard [48]) the distance range is close to 10x (about 40m) that of our experimental system.

3.6.5 Signal-to-Interference Noise Ratio

To facilitate capacity estimation, we measured the signal-to-interference noise ratio $SINR_{\alpha meas}$ in our experimental system.

Let $W_{iON}(x, y)$ and $W_{iOFF}(x, y)$ represent the intensity of a pixel from a white block at location (x, y) on the camera image where the lights were ON and OFF respectively, and i ($i = 1, 2 \ldots 100$) being the index of the image in the dataset (similarly, $B_{iON}(x, y)$ and $B_{iOFF}(x, y)$ represent pixel intensities from a black block). Let $SINR_W$ denote the signal to interference noise ratio for the white pixel and $SINR_B$ for the black, then

$$
\begin{aligned}
SINR_{\alpha meas} &= \frac{1}{2} \sum \left(\frac{SINR_W}{||W||} + \frac{SINR_B}{||B||} \right) \\
SINR_W &= \gamma_{1m} \frac{s(W)}{k(B) + n(W)} + \gamma_{2m} \frac{s(W)}{n(W)} \\
SINR_B &= \gamma_{1m} \frac{s(B)}{k(W) + n(B)} + \gamma_{2m} \frac{s(B)}{n(B)} \\
s(W) &= \frac{1}{100} \sum_{i=1}^{100} \sum_{x,y} (\alpha_m W_{iON}(x, y))^2 \\
k(B) &= \frac{1}{100} \sum_{i=1}^{100} \sum_{x',y'} (1 - \alpha_m)(B_{iOFF}(x', y'))^2 \\
n(W) &= \frac{1}{100} \sum_{i=1}^{100} \sum_{x,y} (W_{iON}(x, y) - W_{iOFF}(x, y))^2
\end{aligned}
\tag{3.8}
$$

where $(x', y') \neq (x, y)$, $||W||$ and $||B||$ represent the total number of white and black blocks respectively. γ_{1m} and γ_{2m} represent the measured number of pixels on the boundary and non-boundary blocks of the imaged block respectively.

We plot $SINR_{\alpha meas}$ versus α, along with the analytical $SINR_\alpha$ from equation (3.4), in Figure 3.7 (a). We can observe from that our SINR measurements are in close agreement with our model (maximum error margin of 1.5dB). We plot the per-block measured SINR $SINR_{blk}(\alpha, B)$ using $SINR_{\alpha meas}$) versus α for different block-sizes B in Fig. 3.6 (c).

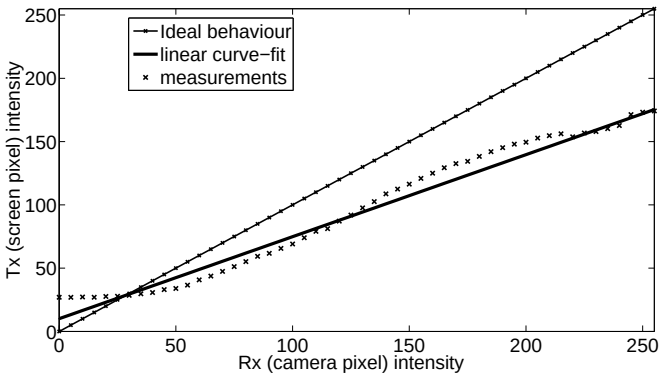


Figure 3.10: Screen-Camera pixel-intensity mapping

Parameter	**Value**
Cam pixel side-length $s_{cam}[\mu\text{m}]$	65
Cam focal length f_{cam} $[\times s_{cam}]$	1573
Screen pixel side-length $s_t[\text{mm}]$	0.248
Principal point (o_x, o_y)	(960.1,539.2)
Noise-variance σ_n^2	101.28
Lens-blur variance σ_{blur}^2 $[\times s_{cam}^2]$	0.25
$\|\|R_s\|\|$ $(=\|\|R_{cam}\|\|)$ [pixels]	1920×1080
Focal-distance $d_f[\text{m}]$	0.39

Table 3.2: Table of screen, camera and measured parameters

We can infer from Fig. 3.6 (c) that, larger the block higher is the the per-block SINR. We can also observe that for a block-size $B = 1$, though it provides large number of parallel channels for multiplexing, the signal energy on each channel is much lower than the noise level, even for medium values of α. In this case, additional signal processing is necessary at the receiver can help decode the low SINR signal with minimal errors. In general, the size of blocks becomes a primary design choice as it affects SINR performance.

3.6.6 Noise Measurement

We empirically measured noise power for SINR computation, to aid analytical capacity estimation. The experiment dataset for this analysis consisted of 200 continuous camera

snapshots of the LCD screen at 2m (and perfect alignment), displaying gray-level intensities from 0-255 in steps of 5 (total 52 sets). Based on our measurements we realized that the intensity mapping between screen and camera can be linear approximated(as shown in Fig. 3.10) and can be numerically expressed as $g(x) = 0.6481x + 10.06$ where $x = 0, 1, \ldots 255$, and the constant 10.06 accounts for the deterministic DC noise in the pixel. The factor 0.6481 can be treated as the path loss factor analogous to RF. As mentioned earlier, the AWGN noise from the background manifests as the temporal variance in the pixel intensity. We compute the noise energy per pixel in our LCD screen- tablet camera channel, using the mean-variance ($\hat{var}(g(x)$: averaged over 52 samples) of the intensity mapping between the screen's actual intensity and the measured intensity on the camera pixel as, $\sigma_n^2 = 10.06^2 + \hat{var}(g(x)) = 101.28$.

3.7 Related Work

Camera based communication using screen transmitters is an example of visual MIMO communication where camera is used as a receiver for information transmitted from arrays of light emitting elements of display screens. In chapter 1 the capacity of a camera channel was estimated by treating the transmitter light emitting array and the camera are perfectly aligned. The channel is considered as an analog communications channel where the signal at the receiver is the sampled photocurrents from each image pixel, and do not take into account the quantization limitations in the camera.

The LCD screen-camera channel capacity estimates [**?**] were based on a water-filing algorithm assuming the camera channel can be equalized to encounter the effects of spatial distortions. But the model and the prototype were designed for a fixed distance of 2m between the screen and camera and did not study the effects of perspective on the estimated capacity and throughputs achieved. Perspective distortion has been studied by the imaging community previously [49, 50], but the fact that the camera is a part of a wireless communication channel (captured object is the light source itself) presents a new domain of challenge for applying imaging models in analyzing communication channels.

Recent research has also seen interest in using cameras to retrieve information from screens [46, 47, 51, 52]. These applications use specific receiver processing schemes to combat visual distortions. For example, PixNet [46] proposes to use OFDM modulation to combat the effect of perspective distortion on images by inverse filtering on the estimated channel, and using forward error correction. COBRA [47] proposes to leverage from encoding on the color channels to achieve throughput gains for smartphone screen-camera communication, but at very short distances (22cm). The fact that several prototypes have been constructed reveals that screen-camera communication is gaining large momentum. However, a representative model and the understanding of information capacity bounds in such screen-camera communications has been an open question. Our work discussed in this chapter addressed this problem.

3.8 Conclusion

In this chapter, we discussed the screen-camera communication application for visual MIMO, where cameras could be used as receivers for data transmitted in the form of time-varying 2D barcodes from display screens. We modeled a screen-camera channel using camera projection theory, which addressed visual channel perspective distortions in more detail than prior works. We discussed and modeled the effect of perspective distortion on the information capacity of screen-camera communications, and validated the same through calibration experiments. Our capacity estimates indicated that, even with the frame-rate limitations in off-the-shelf mobile cameras, data-rates of the order of hundreds of kbps to Mbps is possible even when the 2D barcode from the screen images onto only a small portion of the camera image. While these bounds are much less than the ideal (for example, 8bits/pixel×8Mpixel/frame×2fps for the tablet camera we experimented with), Our findings indicated that camera communications is still promising for medium sized data-transfer or even streaming applications; such as downloading a file from a smartphone screen or streaming a movie from a large display wall. Our estimates indicate that current prototypes have only achieved less than half their capacity, which means that designing efficient techniques to address perspective distortions is still an open problem for building high-data rate camera communications.

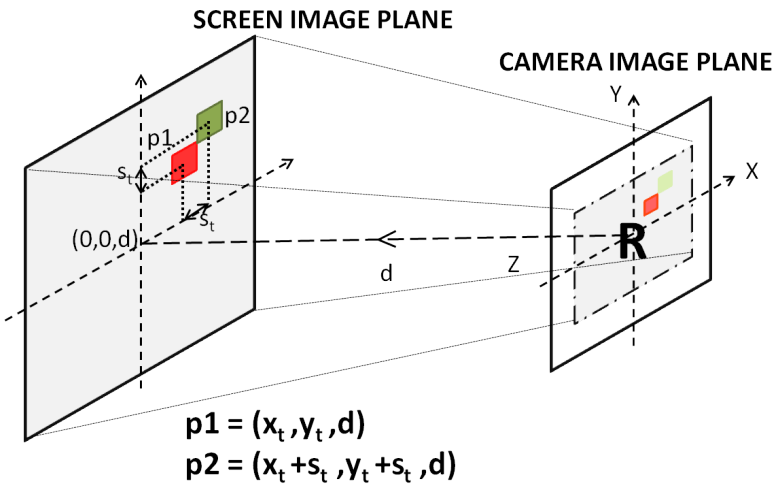


Figure 3.11: Illustration Showing the Screen and Camera Image Axis (observe that, rotation about Z axis will not cause pixel distortion)

3.9 Appendix: Derivation For Perspective Scaling Factor α_p Using Camera Projection Theory

Consider a point $[X_w, Y_w, Z_w]^T$ in world 3D space coordinates with respect to the camera image axis.The camera image 2D coordinates $[x, y]^T$ are given as,

$$\begin{bmatrix} x & y & 1 \end{bmatrix}^T = \mathbf{C} \begin{bmatrix} \mathbf{R} & \mathbf{T} \end{bmatrix} \begin{bmatrix} X_w & Y_w & Z_w \end{bmatrix}^T \tag{3.9}$$

where T denotes transpose operation, $\mathbf{C}$, $\mathbf{R}$, $\mathbf{T}$ are the camera calibration matrix, rotation matrix and translation vector respectively. Camera calibration matrix $\mathbf{C}$ accounts for the projection and scaling of the coordinates in the image ((o_x, o_y) is image center). $\mathbf{R}$ is the rotation matrix that accounts for the 3-tuple rotation angle $(\theta_x, \theta_y, \theta_z)$. and $\mathbf{T}$ accounts for the translation between the world coordinate and the camera axis. If $c\theta = \cos\theta$, $s\theta = \sin\theta$ then,

$$\mathbf{R} = \begin{bmatrix} c\theta_z & -s\theta_z & 0 \\ s\theta_z & c\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} c\theta_y & 0 & s\theta_y \\ 0 & 1 & 0 \\ -s\theta_y & 0 & c\theta_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & c\theta_x & -s\theta_x \\ 0 & s\theta_x & c\theta_x \end{bmatrix} \tag{3.10}$$

$$\mathbf{C} = \begin{bmatrix} \frac{f}{s_{cam}} & 0 & o_x \\ 0 & \frac{f}{s_{cam}} & o_y \\ 0 & 0 & 1 \end{bmatrix} \quad \mathbf{T} = \begin{bmatrix} T_w^x \\ T_w^y \\ T_w^z \end{bmatrix} \tag{3.11}$$

Consider two adjacent pixels p1 and p2 (of side-length s_t) of the screen transmitter situated at distance d from the camera, as shown in Figure 3.11. Let x_t, y_t denote the distance of pixel p1 from the screen's center in X and Y dimensions respectively. Then using camera projection matrix equation from equation 3.9, the distortion in each pixel, $\alpha_{(x_t,y_t)}(x,y)$ can be derived as,

$$\begin{bmatrix} x_{p1} \\ y_{p1} \\ 1 \end{bmatrix} = \mathbf{C}[\mathbf{R}\ \mathbf{T}] \begin{bmatrix} x_t \\ y_t \\ d \end{bmatrix} \quad \begin{bmatrix} x_{p2} \\ y_{p2} \\ 1 \end{bmatrix} = \mathbf{C}[\mathbf{R}\ \mathbf{T}] \begin{bmatrix} x_t + s_t \\ y_t + s_t \\ d \end{bmatrix} \tag{3.12}$$

$$\begin{aligned} \alpha_{(x_t,y_t)}(x,y) &= |x_{p2} - x_{p1}| \times |y_{p2} - y_{p1}| \ \ \forall (x,y) \in \mathbb{R} \\ &= 0, \quad otherwise \end{aligned} \tag{3.13}$$

$$\begin{aligned} \alpha_{(x_t,y_t)}(x,y) = s_t \frac{\frac{f_{cam}}{s_{cam}}(c\theta_y + s\theta_x s\theta_y) + o_x(s\theta_y - s\theta_x c\theta_y)}{x_t s\theta_y - y_t s\theta_x c\theta_y + c\theta_x c\theta_y d} \\ \times s_t \frac{\frac{f_{cam}}{s_{cam}}(c\theta_y) + o_y(s\theta_y - s\theta_x c\theta_y)}{x_t s\theta_y - y_t s\theta_x c\theta_y + c\theta_x c\theta_y d} \end{aligned} \tag{3.14}$$

where $|.|$ denotes the absolute value. $\mathbb{R}$ denotes the set of camera pixels corresponding to the screen's projected image. We assumed that, s_t (order of microns) ¡¡ d (order of cm or m) in our derivation.

Using equation3.14 the average distortion factor α_p can be determined as,

$$\alpha_p = \frac{1}{||R_s||} \sum_{\forall (x_t,y_t)} \frac{1}{||R_{cam}||} \sum_{\forall (x,y)} \alpha_{(x_t,y_t)}(x,y) \tag{3.15}$$

where $||R_s||$, $||R_{cam}||$ are the screen and camera resolutions respectively.

Chapter 4

Throughput Gains by Adapting to Visual Perspectives

Visual MIMO promises to achieve higher information capacity [13] than conventional optical wireless systems that use photodiode receivers especially in mobile settings where ranges greater than tens of meters are required. Due to negligible multipath fading in optical channels the data-rates achievable (i.e., the degree of multiplexing) in visual MIMO depend primarily on the distortions in the visual channel rather than multipath fading, unlike RF. These distortions are typically observed as distortions in the size and shape of the image, partial visibility of an image and even interference between images of two different transmitter elements in the scene due to perspective projection onto the camera sensor and image blurring. In mobile settings, the quality of the visual MIMO link varies significantly with the variation in these distortions which depends on the camera receiver perspective.

4.0.1 Rate Adaptation in visual MIMO

The characteristics of the visual links suggest that the throughput of visual MIMO links can be significantly improved through rate adaptation techniques, which adapt the transmission scheme to the receiver perspective. Particularly in a vehicular setting with front and rear facing visual MIMO transceivers the receiver could provide feedback and make rate adaptation possible. Rate adaptation has, of course, been the focus of extensive study for RF communication systems (e.g., [53–55]). The visual MIMO rate adaptation challenge differs in that (i) the primary challenge lies in MIMO mode adaptation, and (ii) visual MIMO modes present a more complex set of choices than RF rate adaptation algorithms have explored. Mode adaptation is the more significant problem in visual MIMO, since visual MIMO transmitters can employ a much larger

number of transmitter elements than typical RF MIMO systems (due to their operation in the optical spectrum). Selection of the mode requires more complex decisions than the typical rate-up or rate-down decisions of many RF rate adaptation algorithms because the adaptation algorithm has to choose a perspective-appropriate subset of transmitter elements for multiplexing. This subset has to be chosen such that all of the elements are visible to the camera (within field-of-view and not occluded) and so that the transmitter elements do not interfere with each other. Interference between transmitter elements typically occurs when distance increases and the images of the elements start to overlap in the camera view.

To address these challenges, we first define a set of visual MIMO transmission modes for an $N = N_r \times N_c$ LED array transmitter and develops adaptation algorithms to switch to perspective-appropriate modes. The scheme uses packet error feedback to choose the appropriate set of LEDs both over changing distance and changing partial visibility conditions. To identify the set of LEDs suitable for multiplexing, we propose a probing scheme that uses certain spatial patterns and a block CRC scheme that uses separate CRCs for the blocks of information sent from each transmitter element. Using trace-based simulations, we compare their performance with a baseline solution that uses an exhaustive probing search through all LED elements. The simulations are based on a car-following video sequence where the car brake light LEDs are assumed to be the transmitter elements.

4.1 Perspective Dependent Data Rates

In Visual MIMO, the achievable data rate depends largely on receiver perspective. In RF MIMO communication systems, multipath fading can lead to independent parallel channels between antenna pairs. This allows multiplexing of information over these independent channels. With N independent channels used for multiplexing, N symbols can be transmitted simultaneously, leading to an N-fold gain in data rate. Although multipath fading is negligible in the optical spectrum considered here, independent parallel channels also exist in visual MIMO. Consider an ideal full frontal view onto a light emitting array at close distance. The light from different transmitter elements will

fall onto different pixels in the camera image. These pixels can be independently read out, which allows the same multiplexing of information across different transmitter-pixel pairs. In this ideal case, the Shannon capacity for the visual MIMO system with multiplexing can be characterized as in RF MIMO by $C_m = NWlog_2(1+\gamma)$, where W is the sampling rate (frame-rate of the camera) and γ is the signal-to-noise ration (SNR) in a single LED-camera communication system, as discussed in earlier work [13]. This assume that SNR differences from LED to LED are negligible. As in RF MIMO, operation in a diversity mode is also possible. In this mode the same bits are signaled on all (or a subset) of transmitter LEDs. This leads to a stronger signal at the receiver and usually less errors. Note that this is also possible when LEDs are blended together in the image, the signals from multiple LEDs will simply be combined on the receiver pixels. This leads to a capacity of $C_d = Wlog_2(1+N_d\gamma)$, where N_d denotes the number of LEDs transmitting in this diversity mode. The key difference to RF MIMO lies in larger N and very different channel distortions introduced by the optical channel.

4.1.1 Modeling Channel Distortions

In practice, the availability of parallel channel will be affected by visibility issues, perspective distortions, and lens blur.

Visibility: Like any other optical wireless system, visual MIMO requires line-of-sight. An outage will generally occur when none of the transmitter elements are directly visible in the camera image. Only rarely will reflections of the transmitter image be strong enough to be detected by the receiver. A key difference of visual MIMO systems is, however, that only *some* of the transmitter elements may be visible. This can occur when random objects partially obstruct the line of visibility between the camera and the transmitter LEDs. It can also occur when the transmitter is only partially within the field of view of the camera or due to whether effect such as snow flakes and rain drops. Such partial visibility means, that fewer parallel channels are available and the maximum achievable gain will be degraded. We model such visibility issues through an index function $V(n)$, which for each LED $n \in 1 \ldots N$ takes a value 1 when the LED is visible or 0 when it is obscured. The instantaneous multiplexing and diversity

capacities C_m and C_d can then be obtained by replacing the total number of LEDs N with the number of visible LEDs $\sum V(n)$. Clearly, visibility often changes over time. Modeling such visibility changes is beyond the scope of this article.

Perspective Distortion: Changes in viewing angle or distance lead to perspective distortions that can also affect the availability of such independent transmitter-pixel channels. Consider again the full-frontal view onto a transmitter array, but now from larger distance. As distance increases, the image of the transmitter will become smaller. Eventually, light from multiple transmitter LEDs will shine onto the same pixels. At this point the light from these transmitters can no longer be independently read out and the achievable multiplexing gain is again reduced. With changes in viewing angle, the image of the transmitter will be skewed. This can lead to situations where part of the transmitter LEDs shine onto the same pixels, while other transmitter LEDs can still be independently received.

Given the camera parameters as well as the location of the transmitters and camera in 3D space, perspective projection analysis [7] can be used to determine which pixels detect light from which transmitters. For simplicity, let us focus here on the effect of distance. Given a fixed-focal length f of the camera, a spatial distance α between the centers of two adjacent LEDs, and the distance to the camera d, we can calculate the separation of the LEDs on the camera image plane using projection. To be able to independently read the signal from two LEDs, let us assume that a minimum image separation η is required. Thus, multiplexing over all N parallel channels is only possible for distances d below the threshold

$$d^* = \frac{f\alpha}{\eta}$$

. In practice, α and η are likely to be fixed for a visual MIMO system, since it will be difficult to dynamically increase the spacing between LEDs or improve the resolution of the camera. It is possible, however, to indirectly modify α by leaving some LEDs unused. This effectively increase the separation between LEDs in use but decreases the multiplexing gain.

Lens Blur: In addition to perspective distortions, lens blur can lead to blending of the images from two different transmitters. The amount of blur in a camera image is a characteristic of the camera lens and specific to the type of lens used in the system. Such blur is often modeled with a Gaussian blur filter. That is, a (blurred) image Z_{im} as the output of a Gaussian blur filter whose input is an ideal image Z_{ideal}. $Z_{im} = Z_{ideal} * g_{blur}$ where $'*'$ represents a 2D-convolution operation and g_{blur} is a 2D-Gaussian function with zero mean and standard deviation σ_{blur} measured using experiments [7]. In this paper, we will assume that the Gaussian blur from two LEDs can be separated an independently read out, if the distance between the centers is greater than the full-width-at-half-maximum (FWHM) of the Gaussian blur function. FWHM is often used as a parameter for image resolution in analyzing fine detailed astronomical and medical images [11, 12]. That means, we define the minimum necessary separation in the image plane η as follows.

$$\eta = 2\sqrt{2ln2}\sigma_{blur} \tag{4.1}$$

The rate adaptation problem then is to choose transmission modes that exploit the available parallel channels while keeping the error rate low. Multiplexing across more transmitter LEDs will lead to higher data rates, but including an LED that is occluded in the image, for example, would lead to bit errors. We will further discuss an analyze different possible transmission modes next.

4.1.2 Transmission Modes

A transmission mode is a certain assignment of multiplexing and diversity functions to the set of LEDs. In one mode, which we refer to as full multiplexing, bits are multiplexed over all LEDs. In another mode, all LEDs would be used to transmit the same bits. We refer to this as full diversity mode. In between these extremes, lie many other possibilities where only subsets of LEDs are used for multiplexing or some subsets of LEDs are grouped for diversity operation. We therefore define a transmission mode as a a set of non-overlapping subsets from ..., wherein the LEDs in each subset transmit the same bits and information is multiplexed over the different subsets.

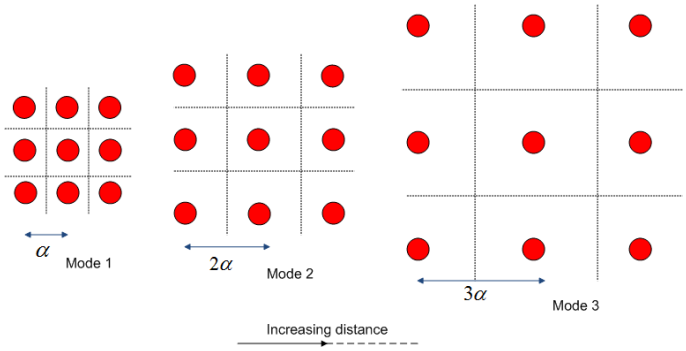


Figure 4.1: Ideal LED array configuration adjustment

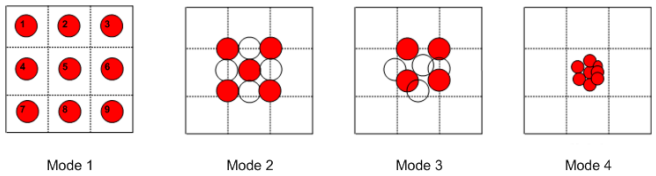


Figure 4.2: Illustration of 3×3 LED array modes for Alternate-LED scheme

As an example, let us discuss some possible modes that can be obtained on a 3×3 transmitter array by choosing subsets of LEDs for multiplexing. Assume that each LED is separated from the next LED in the same row and column by α units. Recall that the full multiplexing mode (mode 1 in Fig. 4.2) can be used only up to a critical distance of d^* and would provide a multiplexing gain of 9. If we now consider mode 2, which leaves LEDs $\{(1,2),(2,1),(2,3),(3,2)\}$ unused, the spatial separation between active LEDs increases to $\alpha\sqrt{(2)}$. This increases the maximum distance to $\sqrt{2}d^*$, albeit at a reduced multiplexing gain of 5. In mode 3, we also switch off LED $(2,2)$, which allows communication for all $d \leq 2d^*$. The system can multiplex over the remaining LEDs $\{(1,1),(1,3),(3,1),(3,3)\}$, yielding a multiplexing gain of 4. The largest range is provided by the full diversity mode.

Other modes can be required to address visibility of the transmitter to the camera. Weather conditions such as fog, rain or snow can significantly reduce the resolvability due to occlusions over time and blurring. For example, if the right half of the LED array was obscured, a multiplexing mode should only include LEDs from the left half. The resolvability is also reduced when the camera is at an angle to the transmitter. One possible mode to address the resulting skewed images is to use a combination

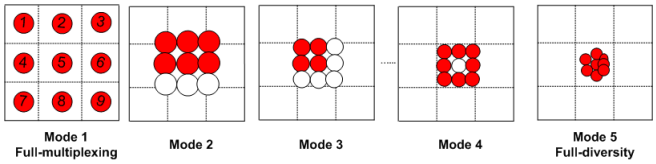


Figure 4.3: Illustration of 3×3 LED array modes for Grouping scheme

modes	(d_{min}, d_{max})	(N_m, N_d)	***mode*_Rate**
1	$(0, d^*]$	(9,1)	$C_{mimo}(d^*)$
2	$(d^*, \sqrt{2}d^*]$	(5,1)	$C_{mimo}(\sqrt{2}d^*)$
3	$(\sqrt{2}d^*, 2d^*]$	(4,1)	$C_{mimo}(2d^*)$
4	$(2d^*, d_{max}]$	(1,9)	$C_{mimo}(d_{max})$

Table 4.1: Modes and rates for N = 3×3 LED array

of multiplexing and diversity where a group of LEDs could coordinate to attempt to provide sufficient brightness for a particular bit, but individual groups could be spaced sufficiently far apart to reduce the chances of blur among the groups. An example of one such grouping is shown in Fig. 4.3 for a 3×3 array. In the vehicular application context, we expect that visibility and distance distortions are more prominent than such angular distortions and the remainder of the paper will focus on these.

4.1.3 The Rate Adaptation Problem and Error Model

Due to the large number of possible transmission modes, the visual MIMO rate adaptation problem lies in efficiently choosing a transmission mode that maximizes throughput. We assume an on-off-keying communication system where feedback in the form of acknowledgments is available. The feedback channel could be realized through a reverse visual MIMO link.

We base our design and simulations on the following packet error model. Recall that for an independent and identically distributed (i.i.d) stream of bits framed into L bit packet sequences the packet error rate (PER) is given as $PER = 1 - (1 - P_e)^L$, where P_e is the bit error probability. A received packet is erroneous if at least one bit in the packet or equivalently one LED is in error. Bit errors may be caused due the AWGN background light noise and also due to visual distortions we discussed. In this context, we consider that LEDs will be in error when their centers cannot be resolved

from any adjacent LED in the image space. Let us consider a visual MIMO LED array with $m = 1, 2 \ldots M$ multiplexing sets or groups where each set transmits a different bit. Each set or group can be set to transmit similar bits, that is, to use diversity. Let D_m denote the total number of LEDs in each multiplexing group m. Then the packet error ratio can be expressed as,

$$PER = \begin{cases} \frac{1}{M} \sum_{m=1}^{M} 1 - [1 - Q(\sqrt{D_m \gamma(m)})]^L V_e(m) & \text{if} \\ \qquad\qquad min(\alpha_{(im)}) > 2\eta, & \\ 1 & \text{if} o.w \end{cases} \tag{4.2}$$

where $Q(\sqrt{\gamma(m)})$ is the average BER for a single LED in the set m with SNR γ in the AWGN visual MIMO channel and On-Off Keying (OOK) modulation (equivalent to the average BER for OOK in AWGN RF channel [4]), $V_e(m)$ is the visibility factor of the LEDs in set m and takes a value 1 (0) when the LED set is visible (occluded). An LED is termed visible when it projects on the camera image such that it is detectable through image processing techniques at the receiver. In our current design we consider that the LEDs are visible even in case of partial occlusions but then we still account for the fact that the signal strength of the LEDs may be low to still result in a detection error. $\alpha_{(im)}$ is the set of the image separation values (in pixels) between any two multiplexing regions. Given the spatial separation α the separation in the image can be found by perspective projection equations [7] as $\alpha_{(im)} = \frac{f\alpha}{ds}$ (f, d, s are the focal length, distance and pixel side length respectively). η is the FWHM of the Gaussian blur from equation 4.1.

4.2 VMRA-Rate Adaptation Algorithms

In this section we detail our proposed algorithms for our receiver-based rate adaptation protocol VMRA that adapts its transmission data rate over distance and visibility variations in the visual channel. The algorithms use the packet error feedback information

to choose the appropriate set of LEDs both over changing distance and changing partial visibility conditions. In our design the LED array transmitter sends a continuous stream of packets - each appended with a CRC - that are decoded at the camera receiver. Upon each successful packet receipt the receiver sends back an acknowledgment (ACK signal) back to the transmitter over an RF feedback channel. The transmitter then flags the transmission as erroneous based on a packet error ratio computed over a time window of T sec (can be of the order of tens of frame-time),

$$PER = 1 - \frac{(\#\_ACKs\_in\_time\_T)}{\#\_packets\_transmitted\_in\_time\_T} \tag{4.3}$$

The tranmission is termed successful as long as the PER is below a preset threshold PER^* (typically 10-15%). In our protocol the transmitter data rate is adapted to the distance variations in the channel by switching to the perspective-appropriate *mode* using the alternate-LED patterns as described in section 4.1.2. Since the data rates in our system is primarily dependent on the number of multiplexing/diversity LEDs over each iteration of the adaptation, we design our algorithms to output the set of indices of LEDs that can be multiplexed β_m ($|\beta_m| = N_m$) and that can be used for diversity β_d =($|\beta_d| = N_d$). To adapt the transmissions to the visibility variations in the channel each of our algorithms use different techniques to determine the set of visible LEDs over each iteration of the adaptation. Such techniques will be discussed along with a detail description of the algorithms in the following sections.

4.2.1 Exhaustive LED search VMRA

This algorithm uses an elementary approach to find which of the bits (LED) are in error by exhaustively searching over all the LEDs (N of them) in the transmitter array. Each LED is set to transmit a known training sequence (of length m). Upon decoding each LEDs signal over a span of $m \times N$ consecutive image frames the LEDs that are in error are determined. But errors in the LEDs can also be due to the merging of two adjacent LEDs due to the distance variation (*mode* change). To check this, immediately after the exhaustive search of the LEDs two corner LEDs of the visible set are set to switch

ON simultaneously for one frame time. Using the image separation between these two corner LEDs, an estimate of the separation between two adjacent LEDs is calculated using perspective projection theory [7] and based on which the *mode* is estimated (and feedback to the transmitter). The algorithm then sets the transmission rate of the system based on the *mode* and the visible set of LEDs that are not occluded. The algorithm re initiates this search process next only when a transmission error occurs or after a time-out t_{out} sec (set to a large value like 10-20sec). If the system does not see any error until time-out then the algorithm sets the transmission back to highest rate (full-multiplexing mode).

4.2.2 Framing based algorithms

The exhaustive search to detect erroneous LEDs may prove wasteful particularly when the the size of array is very large. In such cases rate adaptation may not perform as fast as it is needed to especially in mobile scenarios. Given the spatial setting of the LED array it may be possible to find erroneous LEDs by framing packet transmissions in a spatially coordinated manner over the array. By coordinating such packet transmissions over space and time it may be possible to 'track'the bit transmissions from packets not only in time but spatially as well. In this aspect, we propose two possible techniques of such spatio-temporal framing,

Bit per LED : The LED transmitter array is set to transmit packets of constant size L bits such that each LED transmits one bit. Each packet contains a C bit CRC for error detection. When is the system uses 'multiplexing'each LED transmits an independent bit from the data packet while when using 'diversity'all the LEDs of the transmit array transmit the same bit. The significance of such a framing is that it is practical and easily implementable.

Block per LED: The data packets are split into blocks of data bits. These blocks are spatially framed such that each LED transmits bits corresponding to individual blocks from a packet. Each block is also appended with a C bit CRC for error detection. Only when the system uses spatial diversity on a set of LEDs then the transmitter frames the packets such that the diversity LEDs will transmit the same bits from the same

block. Though such a configuration may be relatively complex when compared to *Bit per LED* the advantage is that detecting the erroneous LEDs (and hence bit errors) becomes very easy as packet errors can be detected just by indexing the erroneous blocks and mapping it to the corresponding LED indices.

Probe-VMRA

In this section we propose a Probe VMRA algorithm that uses the bit per LED framing technique for packet transmissions. This algorithm uses a unique probe function Probe-Visibility() to detect occlusion in the visual channel by using a smart spatial patterning of bits on each row and column of the array. The spatial patterning for Probevisibility() is such that any square LED array can be reconfigured to transmit similar bits on each spatial quadrant of the array and complement bits on each side of the horizontal and vertical axis at the center of the array. The fundamental idea behind using such a pattern is that, by having a copy of the bit and its complement the detection of bit-flips double efficient. This simplifies the detection of erroneous LEDs locations. Other possible patterns to detect occlusions are to use all ones/zeros or alternating ones/zeros. The issue with such approaches is that, in cases of occlusions in the channel the pixel intensity of a bit depends on the object occluding the camera view. If the object is white in color then the pixel intensity will remain high and a bit 1 is retained as bit 1 thus not flagging an error. Alternating ones and zeros may prove helpful but it may only detect if an occlusion has occurred but may not be possible to exactly reveal the occluded LED positions most of the time. In Fig. 2 we illustrate two practical cases in which such a complement bit pattern can work (shown for a 4×4 array but can be extended to any square array). But we also realize that such a probing may not be always error-free. For example, a false alarm can occur when bit b1 in quadrant A gets corrupted due to stronger background noise. In this case the probing detects an occlusion while actually there is no occlusion but noise. Also, the probing may miss an error such as when all the b1 and b1 bits are corrupted such that b1 is detected as b1 and b1 as b1. In this case the probing returns no-error while actually there may have been an occlusion at four locations. But such occurances are very rare in reality

because it is highly uncommon that an occlusion is of the form that can create exactly complementary effects on different spatial regions of the array. False alarms are also rare because the ambient photon noise is typically uniformly averaged over the detector area. Hence we rule out such possibilities and consider the most typical cases in our algorithm design.

- When a transmission error is declared in the system ($PER > PER^*$), the algorithm first checks for the occlusion by initiating the ProbeVisibility() where the LEDs are set to transmit bits based on the spatial pattern mentioned earlier. The function returns the set of visible LEDs indices V and a probe flag pFLAG (true when occluded).

- If the occlusion is full then the probe function would return $V = \phi$. In other cases, the algorithm increments its mode so as to accommodate for any distance variation.

- If the error was due to both partial occlusions and distance variations the algorithm first sets its mode and then reprobes using ProbeVisibility() in the next iteration.

- If there is no occlusion then the function returns a probeerror pError and the algorithm increments to the next transmission mode and checks for errors in transmission.

- In case of multiplexing modes the algorithm sets the LEDs in the visible set V for multiplexing. In diversity mode the algorithm sets all the LEDs for transmitting similar bits (regardless of whether they are occluded or not). If the algorithm is already in its diversity mode then the algorithm resets itself back to full multiplexing and re initiates the probing for occlusions.

Index-VMRA

Here we present a method that obviates the need for exhaustive search or spatial probing to detect the presence of occlusion. This approach uses the block per LED framing

technique of sending packets in the form of independent blocks of bits for each LED. Since each block is appended with a CRC the system can keep track of the erroneous LEDs on the fly by indexing the erroneous block of each packet during a transmission error and indexing those LEDs into a set E. We denote the set of usable LEDs in a mode as U(mode). Using this approach we propose the Index VMRA algorithm .

- In each iteration, the algorithm first indexes all the erroneous LEDs of the set U(mode) into E.
- If the set E is full (all LEDs used in a particular mode are in error) then the algorithm shifts to the next mode and updates U(mode) = set of usable LEDs in the mode.
- If the set E is not full, then all LEDs in U(mode) that are not in error (U(mode)-E) are indexed into the set I.
- If the transmission is in multiplexing mode then all the LEDs in set I are used for multiplexing. In diversity mode, since all LEDs transmit the same bits it may happen that the CRC bits may be corrupted resulting always in error. In this case we start with using all the LEDs in diversity mode and reduce the set by one (LED in any row r and column c) in each subsequent iterations until the transmission is successful ($PER < PER^*$).
- Since it is possible to determine the LEDs that are erroneous in each iteration, over the adaptation period, the erroneous LEDs are set to transmit training packets (packets containing alternate ON-OFF sequences 101010. . .L bits) to determine if the channel is less noisy. Upon successful reception of these bits the receiver acknowledges by sending back an ACK over the visual MIMO feedback channel and those LEDs are indexed to be used for data transmission again.
- The algorithm reinitiates the adaptation procedure when an error occurs in the transmission ($PER > PER^*$) or when a time-out (tout set to a large value like 10-20sec) occurs. If all transmissions are successful for the period of tout the algorithm steps up the transmission rate to next mode with a higher data rate.

4.3 Performance Evaluation

We evaluate the performance of our VMRA protocol in terms of the average throughput achieved by its candidate algorithms over the distance and visibility variations in the channel and compare it with an oracle solution (referred to as *ideal*, that has the power to adapt over the visible set of LEDs to any type of occlusion and distance variation by using the best *mode* for maximizing the throughput). We then elaborate the adaptation behavior of our two framing based algorithms; *Probe VMRA*, *Index VMRA*. Our evaluation uses a trace-driven simulation using traces of input derived from a realistic vehicle-vehicle communication setting.

4.3.1 Obtaining trace inputs

As our source for our traces, we used the video of a real car on a highway with a 3×3 LED array configuration in its brake-lights. The video was captured using our Basler Pylon piA640 camera fitted onto another car at a frame-rate of 60fps and 640×480 resolution. Short sequences of image frames in the video were analyzed partly manually and partly using software to generate the two dataset traces (a. *distance* and, b. $distance - occlusion$) that were used as our test inputs for simulation. To obtain the *distance* trace we used a basic tracking technique from computer vision [7] to estimate the distance between the LED two brakelights x of the car in each image frame of the video. This inter-brakelight separation in the image (in pixels) was used to compute the distance d between the camera and the car in each frame using perspective projection mapping $d = f\frac{l}{x}$, where f is the calibrated camera focal length and $l = 1.5$m is the typical spatial distance of separation between two brakelights in a car. The $distance - occlusion$ trace was obtained by analyzing a set of frames from the video and manually creating a dataset where the number of distinguishable LEDs visible in each frame was noted. We also noted down if the transmitter was fully visible ($visibility = 1$), fully occluded ($visibility = 0$) or partially visible ($visibility = 0.25$ or 0.5 or 0.75) depending on the image area of the array visible. Based on the number of LEDs distinguishable in each frame, the transmission *modes* were manually estimated

and used as ground-truth. The data samples in both traces were spaced by one frame time (1/60 secs). Figure 4.4 shows a few samples of images that were analyzed.

4.3.2 Simulation Methodology

We simulated the adaptation behavior and computed the performance of our VMRA candidate algorithms in MATLAB for the two types of trace inputs using a common simulation methodology. In the simulation each algorithm set to adapt its transmission rate parameter R_{tx} to the number of visible LEDs and the transmission *mode* based on the PER determined (equation 4.2) in each iteration. Since SNRs in a visual MIMO channel are typically very large [13] we assume the probability of bit error due to AWGN background noise is zero. Thus the PER takes values depending only on the errors due to visibility and/or distance change at any LED (r, c). Hence PER from equation 4.2 reduces to,

$$PER = \begin{cases} 1 & \text{if}\{min(\alpha_{(im)}) < 2\eta\} \; or \; V_e(m) = 0 \\ 0 & \text{if } otherwise \end{cases} \tag{4.4}$$

While detecting error due to visibility is done independent of adaptation (probing or indexing using framing) the error due to distance change is detected by comparing the *mode* determined by the algorithm with that estimated based on distance and a resolvability threshold (η = FWHM of Gaussian blur) using perspective projection theory as,

$$mode = \begin{cases} 1 & \text{if} d \leq \frac{f\alpha}{\eta s} \\ 2 & \text{if } \frac{f\alpha}{\eta} \leq d \leq \sqrt{2}\frac{f\alpha}{\eta s} \\ 3 & \text{if } \sqrt{2}\frac{f\alpha}{\eta} \leq d \leq 2\frac{f\alpha}{\eta s} \\ 4 & \text{if } d > \frac{f\alpha}{\eta s} \end{cases} \tag{4.5}$$

When $PER = 1$ a transmission *error-flag* is raised and the adaptation algorithm is initiated which would set a rate R_{tx} based on the *mode* and the number of visible LEDs

Figure 4.4: Sample video frames analyzed for data trace

specified by the algorithm. The data rates for each transmission *mode* are summarized in the look-up table 4.2 for a 3×3 array configuration. To retain uniformity in all cases we set the rate $R_{tx} = 0$ if the LED array is fully occluded.

In order to understand the behavior of the algorithms we recorded a few parameters output at each iteration, such as, a.the transmission *modes*, b.if transmission *error* flag, c.*ProbeError* flag (only for *Probe VMRA*) , and d.transmission rate R_{tx} set by the algorithm. We then computed the average throughput ρ for each algorithms and for each of the two traces analyzed as,

$$\rho = \frac{T}{B} \sum_{i}^{B/T} R_{tx}(i)(1 - error(i))(1 - ProbeError(i)) \tag{4.6}$$

where B is the time width of the window of data-trace. As *Exhaustive LED search*, *Index VMRA* do not use the probing to detect occlusion we set $ProbeError = 0$ when computing the average throughput for these algorithms.

mode	(d_{min}, d_{max})[m]	**(N_m,N_d)**	**Rate** [kbps]
1	$(0, 17.6]$	(9,1)	1192
2	$(17.6, 24.91]$	(5,1)	543.9
3	$(24.91, 35.2]$	(4,1)	338.54
4	$d > 35.2$	(1,9)	59.24

Table 4.2: Rate choice for each *mode* for N = 3x3 LED array with $\alpha = 2$cm interLED spacing

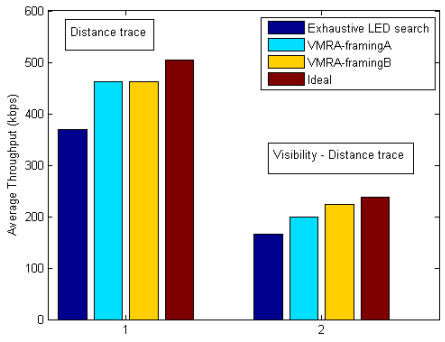


Figure 4.5: Summary of throughput performance

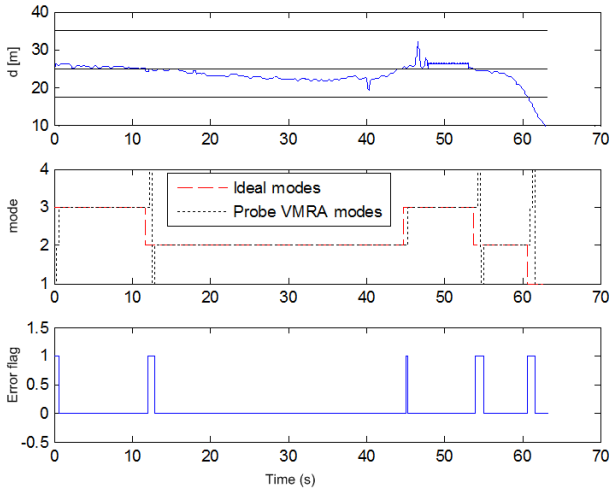


Figure 4.6: Probe VMRA performance over *distance* trace

4.3.3 Trace-driven Evaluation Results

In Fig 4.5 we plot the average throughput ρ from 4.6 for our proposed VMRA algorithms and compare it with the *ideal* performance from the *oracle* solution. We clearly see that our framing-based approaches achieve close to an ideal performance for distance variations as well as visibility variation traces.

We now elaborate on the adaptation behavior of VMRA protocol using the trace-based results for its framing-based algorithms a. Probe VMRA and, b. Index VMRA. We illustrate each algorithm's adaptation behavior over time by plotting the output of each iteration of the algorithm for the *distance* trace and then repeat the same for the $distance - occlusion$ trace. Fig. 4.6 shows the performance of the *Probe VMRA*

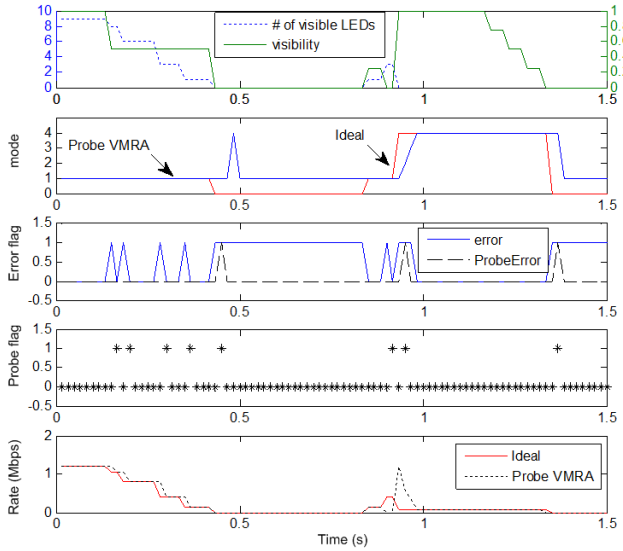


Figure 4.7: Probe VMRA-*distance* − *occlusion* trace

algorithm for the *distance* trace over time. We see that whenever an *error* is declared then the algorithm increments its *mode* until no more *error* is declared. Once the system reaches the mode 4 (diversity) then the algorithm resets back to mode 1 (full-multiplexing). As *Index VMRA* also uses the same approach for adaptation over distance the performance will be the same. In Figure 4.7 and 4.8 we show the performance of the *Probe VMRA* and *Index VMRA* algorithms for the *distance* − *occlusion* trace. Observe that, in the region A where the transmitter is partially visible, the *Probe VMRA* always first initiates the probe for detecting occlusion of the LEDs and only if a *ProbeError* is declared the algorithm changes its *mode*. Thus this algorithm always incurs a 'one-iteration' delay when adapting to the partial-occlusion of the array. The *Index VMRA* on the other hand does not incur any delay or error in detecting the partial-occlusion as the indices of erroneous LEDs are logged over each iteration. Also observe that the outcome of complete occlusion such as in point B ($t \approx 0.4sec$) is that all the packets become erroneous. Since the system has no knowledge if the reason for such packet corruption is complete occlusion or distance variation, both these algorithms first check if the system is in a diversity (*mode* 4) and if not then the algorithm sets the system into a full-multiplexing mode (*mode* 1) by transmitting at the highest

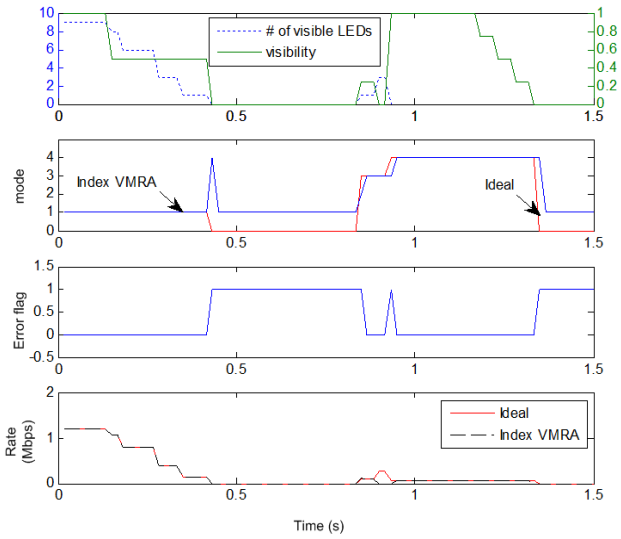


Figure 4.8: Index VMRA-*distance* – *occlusion* trace

rate and then wait for the channel to get better (array being visible).

4.4 Related Work

Rate adaptation protocols have been largely studied, designed, experimented and implemented for the RF channel over the years and especially for 802.11 based networks [53–55]. In IEEE 802.11 networks, the current standard feedback from a receiver to a transmitter is only the presence or absence of an ACK frame. Many rate adaptation schemes also try to rely on physical layer metrics such as signal-to-interference noise ratio (SINR) and bit error rate (BER) to obtain the channel condition [56]. Such schemes typically apply to cellular networks (e.g WiMAX and 3GPP) since cellular networks have a wide range of SINR values [57]. There have also been both simulation and implementation efforts in 802.11 networks to find out what is the best rate along with the choice of spatial diversity or spatial multiplexing for the wide range of channel state among MIMO antennas [58, 59].

Similar to a few RF mechanisms an optical transmitter may also seek to adapt to the link condition change based on a few physical layer parameters. For optical channels, Diana and Kahn [60] investigated how to adjust the parameters of FEC techniques, such as repetition codes and rate-compatible punctured convolutional (RCPC) codes,

for Infra-Red links (IR) based on BER metric. In their paper, Garcia-Zambrana [61] do not adapt the bit-rate directly but seek to enhance the peak-to-average optical power ratio (PAOPR) by inserting the silence period while keeping the average optical transmitted power for FSO links. More recently, Grubor et al. [62] investigated how the power and information bits can be allocated among OFDM subcarriers for throughput maximization. Unlike RF, rate adaptation for visible light wireless links has so far received minimal attention. Most of the approaches have been designed for FSO links and typically use physical layer metrics to quantify the channel conditions over which adaptation is performed. Not much research has been done in designing efficient rate adaptation mechanisms to adapt over the perspective and visibility distortions in visual channels. Applying above techniques such as repetitive codes, RCPC, silence periods to visual MIMO where the camera sampling rates are limited add complexity and significant overheads which can depreciate the effective throughput of the system.

4.5 Conclusion

We probed into the idea that, in visual MIMO the data rate gains from MIMO techniques such as multiplexing and diversity can help achieve throughput gains. In a visual channel such gains are primarily dependent on the perspective distortions in the optical channel rather than multipath fading unlike RF. We discussed how two important factors: (a) distance between the transmitter and receiver, and (b) occlusion of the receiver's view, can govern the quality of the optical link. In this chapter, we proposed a rate adaptation mechanism for visual MIMO using vehicle-vehicle communication as the motivating application example. This work highlighted the necessity to revisit classical rate adaptation methodologies in RF channels when applied to visual MIMO. We proposed a scheme VMRA (Visual MIMO Rate Adaptation) that uses packet error feedback to choose the appropriate set of LEDs both over changing distance and changing partial visibility conditions. We presented three algorithms (*Exhaustive LED search VMRA*, *Probe VMRA*, *Index VMRA*) applicable for our VMRA rate adaptation protocol in an exemplary visual MIMO system that uses LED array transmitter and

camera receiver. The algorithms adapt to distance variations by setting a rate corresponding to the best possible spatial pattern of the elements in the transmitter array (*modes*) at that distance. The *Probe VMRA* and *Index VMRA* use special probe and a block-CRC indexing scheme respectively to detect occlusion efficiently. The *Index VMRA* approach shows the best performance (close to ideal) among the three as the algorithm, unlike the other two, offers an error-free detection of occlusion in the channel and hence adaptation to occlusion incurs minimal overhead.

Chapter 5

Visual MIMO for Low-power Localization

The wide field-of-view of cameras and the directional characteristic of visual links can help achieve mobility in visual MIMO channels. However, the key challenge in realizing mobility in visual channels is the automatic identification of the subjects/objects and positioning in space. In general, the problem can be formalized as a localization problem. Earlier object identification solutions can be broadly categorized into the three approaches: (1) the positioning/tracking approach, (2) the computer vision approach, and (3) the tagging approach. For example, the Wikitude World Browser [63] adopts the first approach – it uses the GPS position and compass together with map information to infer what landmarks a smartphone is pointed at. This approach is generally limited to outdoor use and does not fare well when objects of interest are placed closely together. The second approach – the computer vision approach analyzes the camera footage from a mobile device to recognize objects, which works best with well-known landmarks [64] or previously recognized subjects/objects [65]. The accuracy of this approach degrades, however, as lighting conditions deteriorate, the number of candidate objects/subjects becomes very large, or the objects themselves look very similar (e.g., boxes in a warehouse). The third approach involves tagging objects of interest with a unique identifying code, such as in radio-frequency identification (RFID). RFID allows detecting proximity to objects but orientation (angle) tracking is very challenging in RF channels due to multipath. Directional antennas, though offering more accurate angle information, are too large for wearable devices at typical active RFID wavelengths. Even the angular resolution of a multi-antenna array will be poor if receivers can only be spaced as far apart as the human who is carrying them. This is described by the Airy disc [66], which shows that the smallest angular resolution, θ of a receiving array

with a maximum separation of b between individual receivers is $sin(\theta) = 1.22\frac{\lambda}{b}$. At 900MHz and for head-mounted antenna array spaced by 10cm, this gives an angular resolution greater than 180 deg, and about 90 deg at 2.4GHz. Ultrasound, with its low propagation speed, allows more precise angle-of-arrival estimates but at the cost of increased receiver size (5-10cm) and a significant amount of energy to overcome it's exponential path-loss propagation, as opposed to inverse-square for electro-magnetic waves.

On the other hand, optical solutions, in general are very energy intensive. Cameras are very power-hungry sources and using cameras in a continuous ON operation (as required by most recognition systems) would drain considerable battery power. Today, cameras are more and more being integrated into wearable devices (Google's GLASS or commercial robots and drones). Hence, battery energy including size of such camera devices also need to be considered as important parameters of the object recognition problem. Therefore, designing precise yet low-power object identification remains an open challenge. This thesis addresses this issue through a novel idea that uses a hybrid of two complementary technologies – radio-frequency identification (RFID) and optical signaling. Since we borrow the fundamental idea of using an array receivers at the receiver this work is relevant in the visual MIMO domain. However, as we will discuss further, this work designs a novel camera like device that assists in recognizing objects precisely and which is couple of orders of magnitude more energy efficient than an off-the-shelf CMOS camera.

5.0.1 A Hybrid Radio-Optical Beaconing Approach

We address the low-power recognition problem through our design of a hybrid, radio-optical beacon approach, referred to as *ROP*. ROP, consisting of a radio frequency unit as well as an infrared unit, has the following distinctive advantages. First, infrared beacons can lead to precise angle-of-arrival estimation with a relatively small receiver, due to their small 850nm wavelength. They also do not travel through visual obstructions and are less susceptible to multipath.

In fact, infrared (IR) and visible light signals have been used for positioning (e.g.,

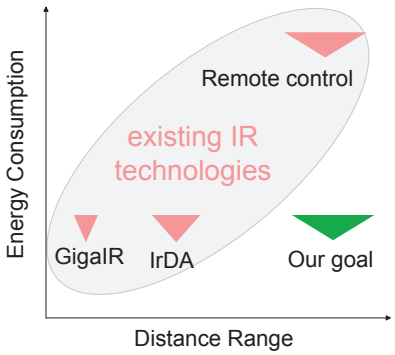


Figure 5.1: Comparison of known IR technologies with our goal for the radio-optical approach in terms of energy consumption, distance range and beam-width (width of the cone in the diagram).

[67–69]), but the high energy consumption has remained a serious impediment to be used for wearable augmented reality applications. As a result, in most prior systems, duty cycles are kept very short and often at least one end of the system (either transmitter or receiver) is not battery operated. Existing IR technologies ([70,71]) typically trade off energy consumption with range and/or beamwidth (angular-range), as illustrated in Figure 5.1. For example, Giga IR [71] can achieve low energy consumption with extremely narrow IR beams, but only within very short distances (i.e., tens of centimeters).

The main advantage of ROP is that it efficiently minimizes energy consumption by timing the infrared beacons using a RF side-channel. Specifically, RF beacons inform the receiver when to expect the IR pulse which allows for using extremely short IR pulses due to tight synchronization between the transmitter and receiver. In addition to reduced energy consumption, short IR pulses also lead to a simplified IR receiver design – instead of requiring an infrared communication receiver (such as in TV remote controls), a synchronized energy detection circuit suffices. Note that this synchronization approach to hybrid radio-optical beaconing significantly differs from existing multi-radio optimization techniques such as low-power wake-up [72] or intelligent switching between radios with different energy profiles [73, 74]. Indeed, existing wake-up techniques are complementary in that the RF link can also be used to wake up the IR beacon when a receiver is detected in the IR range.

Technology	ID	AoA accuracy	Range	Size (order of)	Battery life
IR	encode bits as pulses	high	LOS(<10m)	mm-cm	few days
RFID	ID from data-packets	low	NLOS(< 100m)	few cms	months-years
Ultrasound	require side channel	high	LOS (<14m)	few inches	months
Camera	image recognition	high	LOS (10s of m)	mm-10s cm	1-2 days

Table 5.1: Comparing different positioning technologies. Size of cameras can trade off with speed and image quality

5.1 Applications and Motivation

Recognition based applications typically require that a device can identify the objects within the receiver's view. We assume that the object is attached with a transmitter, or a tag as we call it, while the receiver is attached to the device. Considering that multiple subjects/objects may be within the "view", the task of identifying them demands precise orientation towards these subjects/objects, and association between object positions and identity information. While different kinds of devices can be used, including a smartphone or a robotic vehicle, we use augmented reality glasses as a running example in this paper.

In this section, we first identify a few important requirements for the task of object recognition. Next, we point out that many existing approaches fall short of these requirements, and explain the motivation for our proposed hybrid radio-optical approach.

5.1.1 Requirements of Object Recognition

Precise relative orientation (angle-of-arrival) estimation. Accurate object tracking requires precise estimation of the relative orientation between the object and the receiver, with estimation errors on the order of few degrees. For example, with 1m spacing between shelves in a warehouse and 2m distance between the user and the objects, the angular resolution required to distinguish two shelves is about 15 °; with 10cm

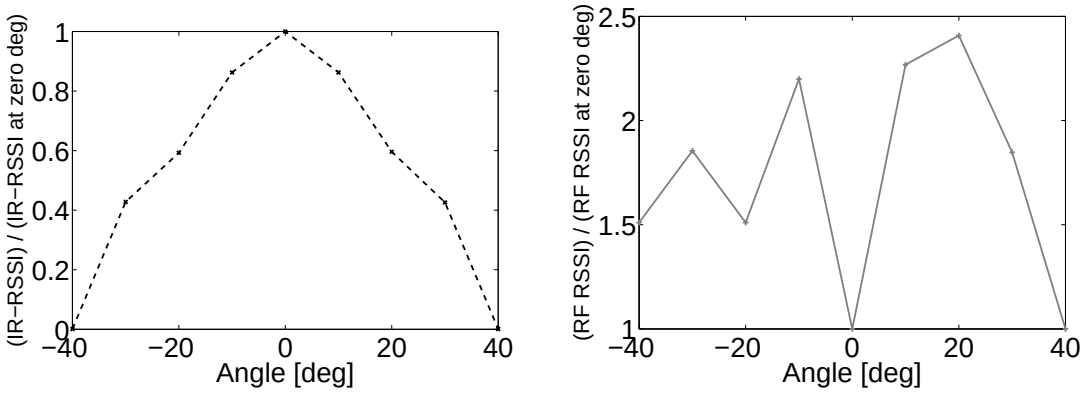
spacing and 3m distance, the angle resolution required becomes 2°. Determining such orientations is similar to determining the angle-of-arrival (AoA) for wireless signals, and therefore we refer to this challenge as AoA estimation. For wearable or head-mounted device such as Google GLASS or augmented reality glasses, precise AoA estimation is particularly important for identifying the objects that are within the wearer's foveal vision—the field-of-view over which the human eye can concentrate to identify an object (which is limited to a $\pm 10^\circ$ range [75]).

Low energy consumption. It is critically important to minimize the energy consumption incurred in object recognition, especially the energy consumption on the transmitter side. For augmented reality applications, we usually have many more transmitters than receivers, and it is not always feasible to periodically recharge/change the transmitter batteries (e.g., transmitters attached to artifacts in a museum, or those attached to shelves in a warehouse). As a result, it is desirable that the transmitter battery lifetime is on the order of several years. On the other hand, we usually have fewer receivers, and it is much easier to charge their batteries (e.g., a pair of glasses).

Small size. Another important requirement is that both transmitters and receivers should be small in form factor, especially the latter as the receiver often needs to be incorporated into a wearable device. For augmented reality glasses, desirable receiver sizes are on the order of about a centimeter, if not smaller.

5.1.2 AoA Estimation Background

We will now examine whether existing technologies that are suited for AoA estimation can meet the key requirements of augmented reality applications. In Table 5.1, we compare such candidate techniques. First, there has been a large body of research on AoA estimation using RF signals [76–78] over the last decade. However AoA using RF is very challenging due to the multipath nature of radio signals— the angle resolution is fundamentally limited to a few radians. Ultrasound signals are good candidates for AoA estimation due to their low propagation speed; enabling precise ranging through accurate time-of-flight estimates. However, ultrasound transducers are costlier than



(a) (b)

Figure 5.2: (a) Infrared RSSI vs. Angle (b) RF RSSI vs. Angle. IR measurements were performed with an IR transmitter LED, a IR Si photodiode, and RF measurements using an omni-directional RFID transmitter and RFID receiver tag at the 907.1MHz frequency (no interfering radios). Receiver was placed 3m away from transmitter in both cases, and at the same well-lit office-room location.

radio antennas and also the receivers have minimum size requirements. Due to its relatively long wavelength, the minimum distance of separation between ultrasound receivers on an array is from few cm to tens of cm, and scales with the number of elements on the array. On the other hand, the highly directional nature of light makes it a viable candidate for accurate AoA estimation. Vision based systems, e.g. using cameras, perform accurate pose-estimation to determine AoA based on preset markers, but cameras are energy intensive. In this paper, we propose a energy efficient usage of the optical spectrum, particularly IR, for positioning.

Advantages of IR (optical) for Positioning. IR, or optical signals in general, due to their highly directional characteristic, can provide robust AoA estimation and ranging, thus are very good candidates for positioning [79, 80]. Figures 5.2 (a) and (b) illustrate the fact that the angular dependency of the received signal strength (RSSI) of IR signal is much more pronounced compared to that of RF. IR is unobtrusive for humans yet it does not travel through obstructions, reducing the likelihood of including objects that are not directly within sight. The signal strength of an IR pulse (beacon) of predetermined duration δ_{ir} received by a photodetector, in terms of the photocurrent generated, can be expressed using the model from Kahn et al. [5] can be expressed as,

$$I_{pd} = \int_{\delta_{ir}} \frac{\gamma P_{led}(t)}{d^2} R_{pd}(\theta) R_{pd}(\phi) dt + \int_{\delta_{ir}} I_n(t) dt, \tag{5.1}$$

where γ is a LED and photodetector specific constant; $P_{led}(t)$ denotes the LED irradiance (in W/sr) or the optical output power of the LED when it is ON at any time t; $R_{pd}(.)$ is the photodetector sensitivity function (normalized s.t. $R_{pd}(0) = 1$) [1]; $I_n(t)$ denotes IR noise current at the receiver and is typically dominated by shot-noise due to background light sources. Due to the nature of the background noise sources in an optical channel, such as indoor ambient-lighting or outdoor sunlight, the noise current changes very slowly with time [5], i.e $I_n(t) \approx I_n$. This indicates that the background noise at any instance can be calibrated by measuring the received photocurrent when the LED is in the OFF state; that is, when $P_{led}(t) = 0$. By performing an inverse operation on equation 5.1, we can estimate the angles θ and ϕ, between the optical ray from a transmitter LED and a reference axis on the receiver using the differential IR signal strength from each detector pair of an array of strategically placed photodetectors. Taking that intuition one step further, in this paper, we propose a object recognition system that adapts the well-known AoA based positioning using IR signals but at much better energy efficiencies than prior works.

5.1.3 Tag Energy Challenge

Though IR signals can give precise AoA estimation and ranging, IR wireless communication is much less energy efficient than RF because it has to overcome much higher ambient noise levels than in the RF spectrum. As an example, a short 10μs IR pulse in our system consumes a similar amount of energy as a 12 byte, 400μs radio packet. Hence, we propose a hybrid approach that offloads not just all communication tasks (conveying identify) to RF and retains IR only for positioning, but it also uses RF signals to provide IR pulse synchronization to the receiver, which allows use of extremely short IR pulses. We will discuss the details next.

---

[1] Note that $R_{pd}(.)$ is typically symmetric along the azimuthal and polar axis for most photodetectors used today.

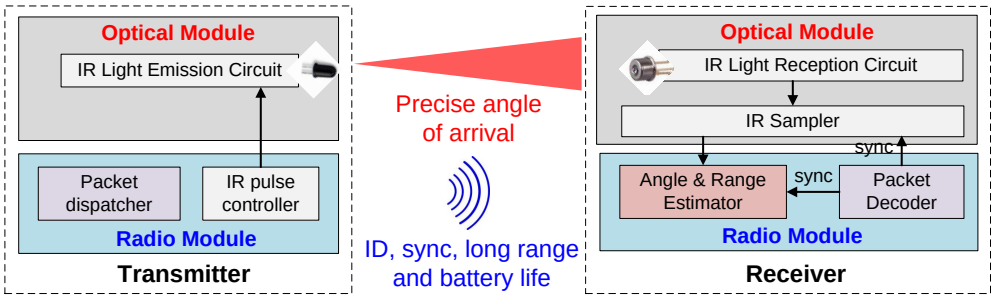


Figure 5.3: ROP system architecture. The IR beacon is used for accurate positioning through AoA, while synchronization and ID communication is through radio.

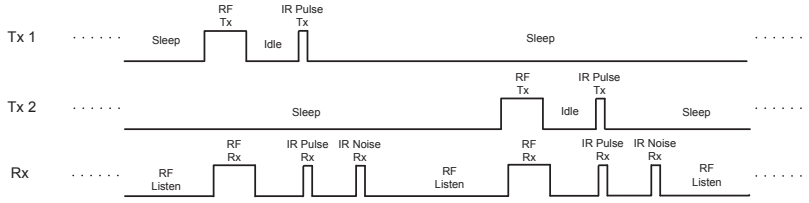


Figure 5.4: Timing diagram of the *paired-beaconing* protocol in ROP over one duty-cycle period (1sec in our prototype) for a 2 transmitter tags example.

5.2 Radio-Optical Beaconing (ROP) System Design

In this paper, we advocate a hybrid radio-optical beaconing approach (ROP) to meet the accuracy, energy, and size requirements of relative orientation tracking. In ROP, the task of orientation tracking involves answering the following three questions: (1) when does the object transmit information, (2) what information does the object transmit, and (3) where is the target. ROP relies on the RF link for the first two questions, and the IR beacon for the third question. Relative orientation tracking enables applications that recognize what object you are looking at. Figure 5.3 illustrates this architecture of ROP. In this section, we will discuss the two key technical aspects of ROP—minimizing energy through a novel IR pulse synchronization technique and positioning through IR signal strength.

5.2.1 Low Power IR Synchronization through Radio Communication

ROP uses a radio-synchronized IR pulse for recognizing (positioning and identification) a tagged object. IR requires high power to overcome the high background illumination

and noise (artificial or solar) in this spectrum. We therefore minimize the transmission period of the IR signal to the point where it can no longer be used for communication purposes but is still detectable for our angle and distance estimation needs. Theoretically, a single short IR pulse with a high peak power, like an optical strobe light, could provide a range[2] of 10m as well as very low *average* energy consumption due to its short duration. The challenge, however, lies in detecting when such a signal is transmitted at the receiver, since the signal lacks the preamble information that is often used for such purposes in communication systems. Adding a preamble would require multiple and possibly longer pulses, which leads to higher energy consumption. High efficiency detection therefore depends on proper timing—by enabling the detector only when the optical pulse is expected, we can eliminate much of the effect of background noise.

ROP addresses this challenge by using RF signals for synchronization. Various ROP protocols can be adopted, and here we discuss a simple protocol, which is referred to as *paired-beaconing*. As illustrated in the timing diagram of this paired-beaconing protocol in Figure 5.4, the transmitter periodically transmits a RF beacon following a suitable communication protocol such as CSMA; immediately following the transmission of a RF beacon, after a very short predetermined delay (known to the transmitter and receiver), an IR pulse is sent out. The receiver uses the end of the radio packet as a reference to synchronize with the incoming IR pulse, and then samples the received signal from the photodiode(s) over the expected pulse duration. It also takes an additional noise measurement after the pulse to calibrate for noise level. The IR pulse itself carries no bits of information—the target's information (e.g., its ID) is included in the preceding radio packet.

By using the RF link to synchronize IR beacons, we significantly cut down the system energy consumption compared to an IR only object recognition system. Note that we could further decrease the system energy consumption by adopting more sophisticated ROP protocols. For example, a protocol could have the transmitter only send out RF beacons until an ACK packet from the receiver is received, following which it sends out

[2]We define range as the maximum distance along line of radiation at which the optical signal can be detected.

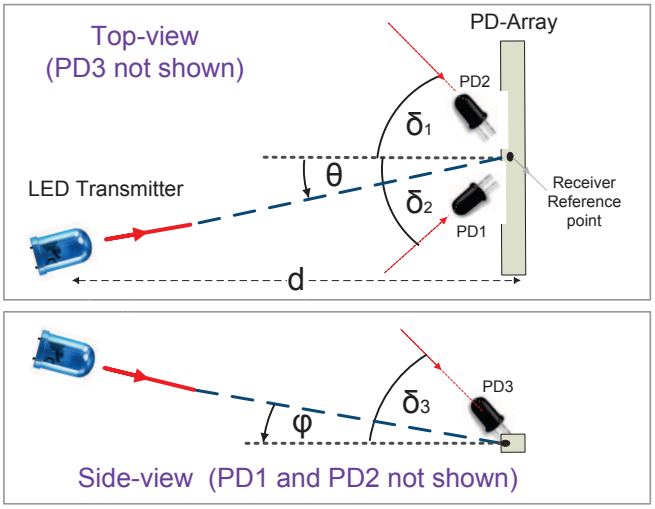


Figure 5.5: Single LED transmitter - three element PD (photodetector) array receiver model ($\delta_1 = \delta_2 = \delta_3 = \delta$)

an IR beacon. Since the energy benefit of such wake-up mechanisms in radio channels has already been studied in the literature (see [72–74]),for the rest of the paper, we will focus on the energy benefits of our proposed radio synchronized IR beaconing mechanism which is significantly different from such wake-up mechanisms.

5.2.2 Positioning Using IR Signal Strength

The ROP receiver design includes a three-element photodetector receiver, sampling IR signals from an IR LED, as shown in Figure 5.5. We define the angles θ and ϕ that we want to estimate as the angle between the receiver surface normal and the vector connecting the transmitter and the depicted reference point in the center of the photodetector array, on the horizontal and vertical planes respectively. We will use the terms *horizontal*, *vertical* to refer to the azimuthal and polar planes respectively. The photodiodes are rotated by an angle δ from the surface normal such that the angle between the LED and the vector in direction of photodiodes 1, 2 and 3 is $\theta - \delta$ and $\theta + \delta$, and $\Phi + \delta$, respectively.

Let I_{h1}, I_{h2}, and I_v represent the noise-subtracted IR signals ($I_{pd} - I_n$, from equation (5.1)) on photodetectors 1, 2 and 3, respectively. We will consider that the photodetector sensitivity is lambertian ($R_{pd}(x) = \cos^n(x)$ where $n \geq 0$, typical for most off-the-shelf photodiodes available today), which can be verified from the datasheet

specifications. Based on equation 5.1, the noise-subtracted receiver photo-current generated due to the light energy accumulated from an LED emitting a constant light output [3] over a duration δ_{ir} is proportional as,

$$\begin{aligned} I_{h1} &\propto \gamma \cos^n(\theta - \delta) \cos^n(\Phi) \\ I_{h2} &\propto \gamma \cos^n(\theta + \delta) \cos^n(\Phi) \\ I_v &\propto \gamma \cos^n(\theta) \cos^n(\Phi + \delta) \end{aligned} \tag{5.2}$$

The position, in terms of the angles and distance, can hence be derived in closed-form using an inverse operation as,

$$\begin{aligned} \theta &= \pm \tan^{-1}\left(\frac{1}{\tan(\delta)}\left|\frac{(\frac{I_{h1}}{I_{h2}})^{1/n} - 1}{(\frac{I_{h1}}{I_{h2}})^{1/n} + 1}\right|\right) \\ \Phi &= \pm tan^{-1}\left(\cot\delta - 2\left(\frac{I_v^{\frac{1}{n}}}{I_{h1}^{\frac{1}{n}} + I_{h2}^{\frac{1}{n}}}\right)\mathrm{cosec}\delta\right) \\ d &= \left(\gamma \frac{\cos^n(\theta_{est} + \delta) + \cos^n(\theta_{est} - \delta)}{(I_{h1} + I_{h2})}\right)^{\frac{1}{2}} \end{aligned} \tag{5.3}$$

where $\pm$ indicate the directions relative to the reference point (left or right, up or down). The scale factor γ can be determined using parameter values obtained from the LED and photodiode datasheets.

5.3 Prototype Design

We have prototyped the ROP tag and a receiver that identifies and estimates relative position (AoA and distance) of a tag in range. We have also incorporated the receiver into a wearable prototype and developed several applications by mounting the receiver circuitry on a glasses together with a RECON Instruments MOD LIVE heads-up-display that runs Android. We focus here on the tag and receiver implementation.

[3] $P_{led}(t)$ in equation 5.1 is a non-zero constant over the δ_{ir} duration, and zero otherwise

5.3.1 Radio-Optical Transmitter Tags

The transmitter tag consists of a RFID module that is used for the radio communication as well as triggering the pulse input to an IR LED. To be detectable at maximum distance the LED has to be operated for maximum light emission. The LED achieves maximum light emission when the current (voltage) across the LED is 1A (2.5V). As illustrated in the tag circuit diagram in Figure 5.6, a MOSFET amplifier and an appropriate series resistor ensured that the current across each LED was maintained at 1A. To maximize range, we used two near-IR LEDs [81] on the prototype tag. A high-energy pulsed LED emission requires a large spike in energy which cannot be achieved if powered by the same power supply of the radio. So we use an independent 9V battery supply for the driving the LEDs and use a capacitor to prevent a sudden large voltage drop when the LEDs are activated. The 9V power supply can be avoided by using a lower voltage battery along with a voltage step-up circuit. Over-driving the LED for maximum range can also be avoided by using multiple high power LEDs at nominal current drive. We reserve such design considerations for future.

The RFID module on the tag contains a CC1100 radio and a MSP430 microprocessor and powered by a CR2032 – 3V lithium coin cell battery. The radio operates at a data rate of 250kbps with MSK modulation and a programmed RF output power of 0 dbm. In each duty cycle, the radio broadcasts a 12-byte packet (4 bytes of preamble, 4 bytes of sync, 1 byte of packet length, 3 bytes of tag id + parity bits), waits for short delay, triggers a 3V pulse for a duration of $\delta_{ir} = 10\mu$s on one general purpose I/O pin connected to the MOSFET gate, and goes back to its sleep mode. The radio wakes up every $\tau = 1$ sec and repeats the transmission. The delay was measured to be at least 500μs: over-the-air packet time of 380μs and 120μs hardware delay.

5.3.2 Radio-Optical Receiver

The front end of the receiver consists of three Silicon photodiodes [82]. Two of them are horizontally spaced by 3 cm and mounted with 40° separation ($\delta = 20°$ symmetrical on each side); the third is placed 20° off (on top) the horizontal plane formed by the

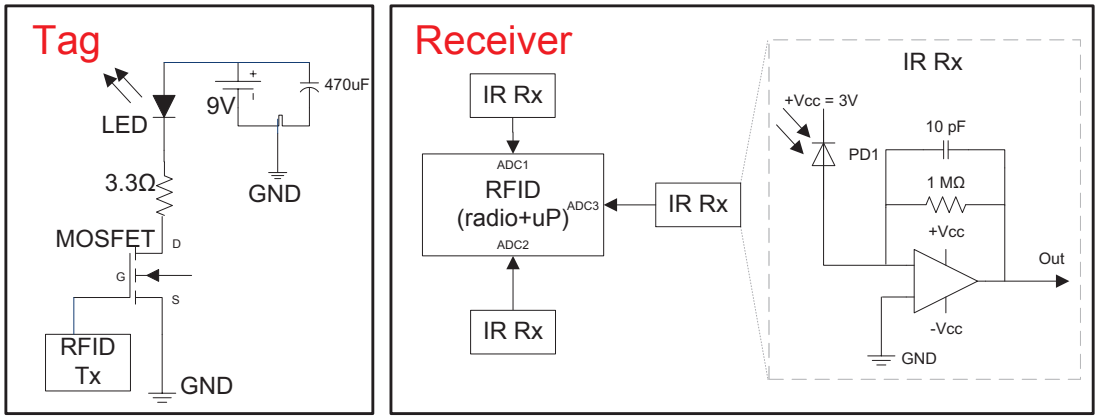


Figure 5.6: Prototype ROP transmitter (tag) and receiver circuit diagrams.

other two. We chose $\delta = 20°$ because it simplifies distance estimation [4] and provides good AoA estimate resolution over an angular-coverage of $\pm 20°$. The angular-coverage of the system can be increased by placing more photodiodes in the receiver array. To amplify the detected signal currents on the photodiodes, we use a opamp (operational amplifier) and choose the resistor and capacitor values in the opamp circuit such that the rise-time (proportional to the time-constant RC) is much less than the IR pulse length, so as to ensure maximum IR light energy accumulation over the pulse detection period at the receiver.

The receiver RFID module contains a CC1100 radio and a MSP430 microprocessor, similar to the radio-optical tag. Each photodiode's analog output from the opamp is wired to each of the three 12-bit analog-to-digital converter (ADC) input pins of the microprocessor. We power the radio using one of the 3V supplies to the opamp (the opamp requires a +Vcc and -Vcc supply). We programmed the radio to stay in always-active receive mode ready for receiving the radio packets and IR beacon. Upon a successful packet reception the signal from the photodiodes are sampled at each ADC, and at a time instance after the end of packet reception – subject to a small hardware delay. The ADC sampling duration is set equal to the length of the IR pulse. The receiver identifies each tag through the unique transmit ID encoded in the radio packet. The sampled ADC voltage readings correspond to the received IR

[4] as the numerator in expression for d in equation 5.3 is almost independent of θ for the LED and photodiode pair that we used

(a) Tags on posters (b) Tags on shelves (c) Tags on objects (d) Tags on objects

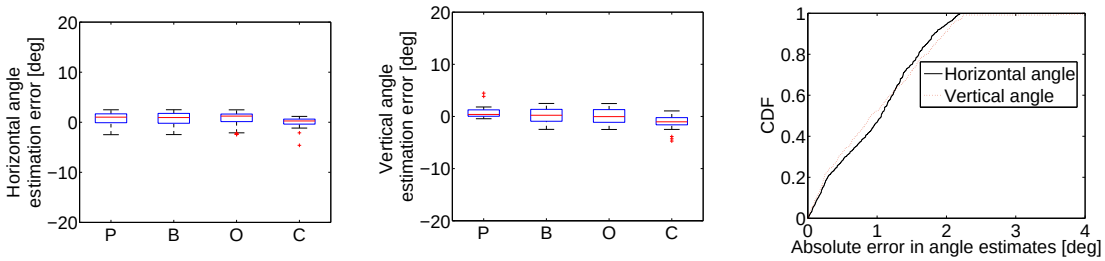
Figure 5.7: (a)-(d) Experimented application scenarios. In all scenarios the tags and the on-looking experimenter faced each other. In the *Office-Room* scenario the experimenter was seated on a chair.

signals; let us denote them as V_{h1}, V_{h2} and V_v. After obtaining the signal samples, the background noise (voltage) is measured by sampling the photodiode outputs after a 60μs delay (10μs of opamp delay plus 50μs pulse fall-time), and for a duration equal to length of IR pulse. Let us denote the noise readings as N_{h1}, N_{h2} and N_v. Since the load resistance is the same for all the voltage readings, the angle and distance are estimated by substituting the numerical values of $V_{h1} - N_{h1}$, $V_{h2} - N_{h2}$ and $V_v - N_v$ values into I_{h1}, I_{h2} and I_v respectively, in equation 5.3.

5.4 Experimental Evaluation

In this section, we present our evaluation results of the ROP system. We have conducted extensive experiments in a well-lit academic laboratory environment and evaluated the performance of ROP in different real-world application settings, primarily in terms of the angle and distance estimation accuracy and battery lifetime. We also evaluate application metrics such as recognition accuracy and latency.

By analyzing more than 15000 data points collected from the experiments, we observe that the median angle estimation error was in the order of 1 deg, for both horizontal and vertical dimensions, and distance error within 40cm. Our power measurements shows that the ROP transmitter consumes around 86μW of power, while the receiver consumes less than half the power required by other existing prototypes.



(a) (b) (c)

Figure 5.8: (a),(b) are horizontal and vertical angle estimation error respectively (P, B, O, C refer to *Posters, Bookshelf, Office-room, Cubicle* scenarios respectively), and (c) shows the aggregate CDF of the angle estimation errors, for four real-world application scenarios

5.4.1 Object Recognition Accuracy and Latency

We have conducted extensive experiments representing four real-world scenarios (shown in Figure 5.7): (i). *Poster* (Figure 5.7(a)), that represents a scenario where users walk from a poster to another while getting background information of each poster through an augmented reality system, (ii). *Bookshelf* (Figure 5.7(b)), that represents a scenario where users desire to locate a certain shelf in a library or warehouse, (iii). *Office-Room* (neat;Figure 5.7(c)), that represents a scenario where a user tries to locate an object in a relatively large and neat office in which objects are spread out, (iv). *Cubicle* (cluttered;Figure 5.7(d)), that represents a scenario where a user searches for items in a cluttered, small space, such as a cubicle or a medicine cabinet.

Experiment Methodology. To facilitate ground-truth angle measurements, we attached a camera [5] (recording video frames at 30fps) fit with an IR lens (will refer to as IR camera from here on) onto the glasses. The reason we chose a camera for ground-truth angle measurements is that, the angle subtended by the light ray with the camera reference axis can be determined accurately using the pixel image coordinate of the light emitter (captured as a white blob by the IR camera). We fit the photodiode array onto the camera such that the reference axis of the photodiode array and camera are the same. This setup avoids errors due to any discrepancy in ground-truth measurements and movement of the array. For manual visual verification, we also fit

[5] a 10μs IR signal integrated over the 33ms frame period (30fps) was detectable by the CMOS sensor of the camera, due to high light energy output from the LED

a smart-phone camera onto a helmet that was worn by the experimenter during the course of experiments.

In each experiment scenario (*Poster, Bookshelf, Neat Office, Cluttered Cubicle*) a total of five transmitters were used, that beaconed an IR pulse of width 10μs every 1 second. A total of 15000 data samples[6] (over 4 hours of experimentation) were collected over multiple trials where, one of the authors, referred to as experimenter, wore the prototype glasses, and performed the following actions in each scenario:

(i) *Poster*: The experimenter read a poster, from a distance of 2m, for a few minutes and moves to the next. Before moving to the next poster, the experimenter would first turn head to look at the subsequent poster from the current location and then walk to it.

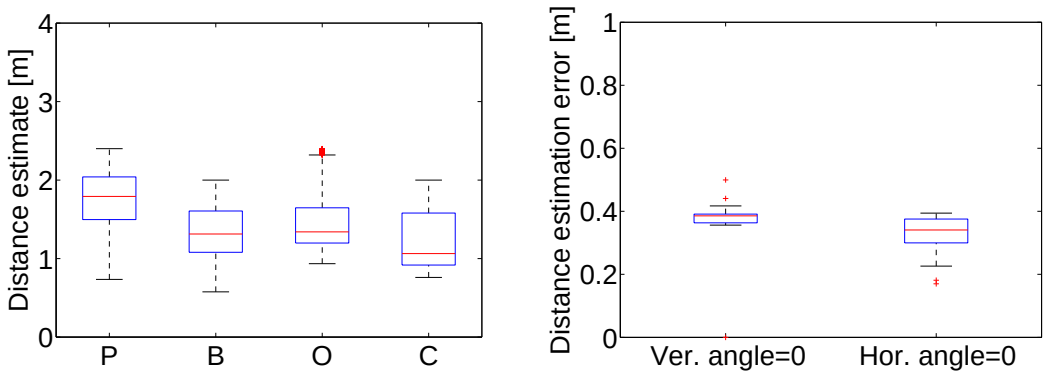
(ii) *Bookshelf*: The experimenter searched to locate a particular bookshelf. Here, the experimenter first tried to locate the shelf (which involved standing at 1.5m from the shelf and looking up or down) and then made slight lateral head-movements to emulate searching for a particular item on that shelf, and then repeated the same exercise for the subsequent tagged shelves.

(iii) *Neat Office*: The experimenter searched for a particular tagged object in the room, looked at it for a few seconds, and did the same for other tagged objects, one after the other. During the course of the experiments, the experimenter was seated on a chair at 1.5m distance along the 0 deg axis facing Tag 2 in Figure 5.7 (c).

(iv) *Cubicle*: The actions in this experiment were the same as in the *Neat Office* scenario, but with the tags placed in a more cluttered space. During the course of the experiment, the experimenter was standing at 1.5m distance along the 0 deg axis of Tag 3 in Figure 5.7 (d).

(a) AoA and Distance Accuracy Results. In Figures 5.8 (a) and (b), we plot the errors in horizontal and vertical angles estimates, respectively. We also plot the cumulative distribution plot of the angle errors in Figure 5.8 (c), and observe that the median error is 1.2° and 80% of the errors are contained within 1.5°, which is closely

[6]All the data, along with timestamps, was collected on a linux laptop with the camera being connected to the laptop through USB



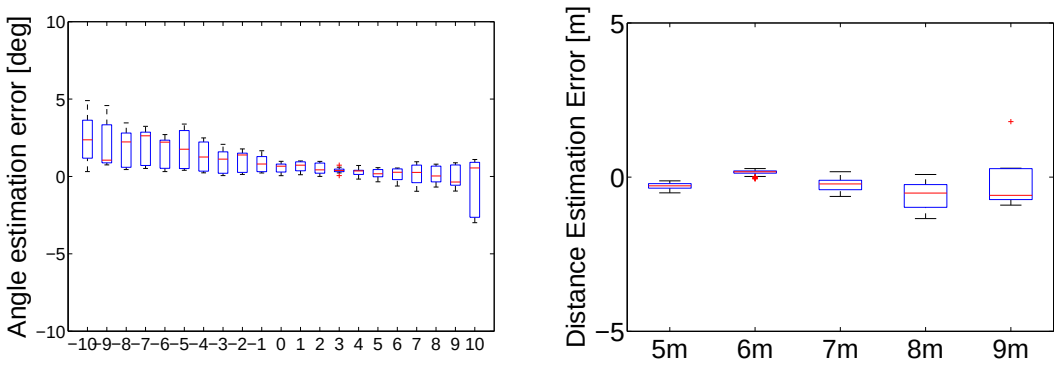
(a) (b)

Figure 5.9: (a) Distance estimation error for four application scenarios (P, B, O, C refer to *Posters, Bookshelf, Office-room, Cubicle* scenarios respectively), (b) Distance estimation error for calibrated head-worn receiver setup

consistent for both horizontal and vertical angles. We believe that an accuracy of 1° is sufficient for many augmented reality applications as discussed in Section 5.1. We note that the angle estimation errors reported here also include the deviations in the ground truth angle measurements due to head movement. We examine this further in section 5.5.

In Figure 5.9(a) we plot the distance estimates from our system (instead of distance estimation errors, since we did not have an accurate measure of ground-truth distance due to the uncontrolled movements in the experiment scenarios). However, we observe from our results that the distance estimates are close to the distances the experimenter maintained during the course of experiments (2m for the scenario(i) and 1.5m for others).

To demonstrate our distance estimation accuracy, we conducted a controlled experiment where the experimenter, wearing the glasses receiver, positioned the head so as to look at one tag. Two sets of data were collected, where in each, one angular dimension (horizontal or vertical) was fixed (to 0 deg) and the other changed; the perpendicular distance between the experimenter and tag was fixed at 3m. We report the distance error estimates from this controlled experiment in Figure 5.9(b) and verify that the median distance error is within 40cm in both cases. We believe that such ranging accuracies are suitable for many AR applications.



(a) Angle estimation error (deg) vs. Horizontal angle (deg)

(b) Distance estimation error (m) vs. Distance (m)

Figure 5.10: (a) Angle estimation error for calibrated (fixed) setup , (b) Distance estimation error for calibrated (fixed) setup

We have also conducted controlled experiments to measure the position estimation accuracy. In this set of experiments, we marked locations on the laboratory floor for ground-truth angle and distance measurements. The measurements spanned -10 to 10 deg in 1 deg spacings on the horizontal ($\theta_{ver} = 0$) and from 5m to 9m in 1m steps. At each marked test points, we positioned the tag and collected 60 consecutive beacon samples. We then repeated the entire procedure 5 times, yielding a total of 300 samples per test point. We performed our evaluations for the tag beaconing period of 1 sec and an IR pulse length of 500μs to maximize range. In this experiment, the receiver glasses and transmitter were both positioned (fixed) on a crate at an equal height of 60cm from the floor. As can be seen from the angle error and the distance error plots in Figure 5.10 (a), (b) respectively, the median angle error (of 1.2 deg) and median distance error (of 40cm) from the application scenario experiment, is indeed very close to the results from the calibrated setup.

(b) Object Recognition Accuracy Results. Note that for augmented reality applications, *the important application-level metric is object recognition accuracy instead of position estimation accuracy.* Each tagged object has a unique object ID and location; if ROP successfully receives and decodes the ID, and estimates its position (both angle and distance) within a preset area centered around the object's true location, it is considered the recognition is successful. The total recognition accuracy is the percentage

of the successful trials with respect to the total trials. Our results indicate that, on an average, a tag within a user's view is successfully recognized with a success rate of 97.5%. The success rates for the *Poster, Bookshelf, Office, Cubicle* scenarios were 97.1%, 98.025%, 97.5%, and 98%, respectively.

Recognition Failure rates. In ROP, false-positive events are primarily triggered due to reflections of the IR signal. Across all four scenarios, the observed false-positive rate is within 2%, among which the *Poster* scenario had the most false-positives due to reflections of the IR signal from the smooth plexiglass surface. In section 5.5 we provide our findings from a simple experiment that characterize the IR reflection level from various common surfaces. On the other hand, false negatives occur when the RF or IR beacon is lost. In our experiments, we observe a false negative rate of 0.5% across all the scenarios. We note that this rate may go up when the tag density increases, but we expect efficient MAC protocols can effectively minimize RF collisions under a reasonable density.

(c) Object Switching Latency. ROP may yield erroneous position estimates when the receiver switches objects suddenly, due to the discrepancy in signal strength on the photodiodes. Temporal averaging will filter out noisy estimates in these situations, but it will require a minimum wait time for the receiver to focus on each transmitter before switching to another.

This object switching latency changes with the beaconing period in ROP as ROP needs a certain number of beacons to identify (with an error margin of $\pm$ 2 deg) the position of a transmitter. Let $\Delta_t = t_1 - t_0$, where t_0 denotes the time instance when tag A goes out-of-view when the user starts to shift head position from tag A to tag B, and t_1 denotes the time instance when tag B is successfully located (within $\pm$ 2 deg). For ground truth of these time instances, we rely on IR camera time-stamps and visual verification from helmet camera feed of the data from our experiments, and define the object switching latency as $\Delta_t^{ROP} - \Delta_t^{ground-truth}$.

We determined the object switching latency of our system, using a total of 100 head-turn events (switch head position from one tag to another) that we obtained from our collective dataset, to be 0.75 seconds on an average and consistent across the four

State	Duration δ [μs]	I_{bat} [mA]	Energy [μJ]
Tag idle	800	2.95	7.08
RF transmit	500	15.52	23.28
IR transmit	12	543	52.78
sleep	998688	0.0007	2.097
Total energy E_{tot}			85.23

Table 5.2: Energy consumption of ROP prototype Tag for a 10μs IR beacon (2 LEDs on tag) and radio transmitting a 12byte packet at 250kbps every 1sec at 1mW (0dbm) output power. Energy = $V_{bat}I_{bat}\delta$, where V_{bat} = 3V for radio module and 8.1V for IR.

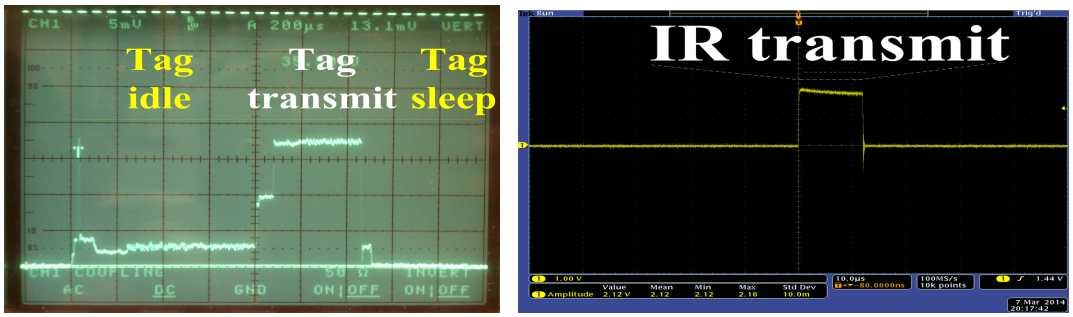
scenarios; for a 1sec beacon period this latency translates to requiring about 2 beacons to accurately recognize the tag when a user makes momentary head movements. The latency can be reduced by choosing a higher tag beaconing rate, but that will cause more battery energy drain (see section 5.4.2).

We believe that the delay of 0.75 sec when a user impulsively shifts head position (as achieved by our prototype) is acceptable for most application settings of ROP. We emphasize that the object switching latency metric considered here is different from the typical *system response time* metric – equal to the time taken by the system to report the output (tag is identified and located) after the input is given (tag transmits). We determined the system response time for our ROP prototype to be 25ms; that includes the tag, receiver, a local network server and an application on the Android phone.

5.4.2 Transmitter Power Consumption

The total energy consumption of the ROP transmitter includes the amount of energy consumed by the three modules: microprocessor, radio, and IR, among which we focus on the latter two modules in this study. In Table 5.2, we report the energy consumption[7] in different states of operation during a 1s beaconing period. We measured the current draw from the battery source in different states of operation – separately for the radio and IR modules as they are powered by independent battery sources; Figures 5.11(a) and (b) show the voltage across a resistor in series with the battery source during

[7] I_{bat} = total current in each state of the tag by integrating corresponding regions of oscilloscope readings from Figure 5.11 (a) for radio, and (b) for IR module



(a) (b)

Figure 5.11: (a) Tag's radio module battery drain (voltage reading is across a 1Ω resistor on an analog oscilloscope), (b) Tag's IR module battery drain (voltage reading is across a 3.9Ω resistor on a digital oscilloscope

normal operation. The 'Idle'state in Table 5.2 includes the transitioning periods from sleep to ON and vice-versa.

Finally, we compute the tag average power consumption as $P_{avg} = \frac{E_{tot}}{\tau}$, which is as low as 85.23μW for a 1 second beaconing period in our prototype.

(a) Comparison With Other Prototypes.

In Figure 5.12 we compare the peak power consumption (product of maximum current draw and supply voltage of battery for active transmission) versus the pulse length of the ROP transmitter, with that of an IR remote control [8] and an ultrasound based positioning prototype (SpiderBat [1]).

We can observe from the area under the curve – that yields the energy, for each technology in Figure 5.12, an IR remote control technology is a less energy-efficient option for fundamental operation of positioning. This is because the IR transmission will have to communicate a packet of bits to replace the RF module in our system, thus keeping the battery on and draining the peak power for a longer duration (due to the need for communicating more pulses where each pulse translates to communicating one bit). While the ultrasound transmission seems to be as energy efficient as ours, however, the reception is about 2.5x less efficient than our system (as we will discuss in the next subsection).

[8] we measured the IR pulse period is 10ms and peak current draw is 50mA from a 3V (two alkaline AAA batteries) supply for a TV remote control. We interpolate the effective pulse-period to be 1ms for transmitting 13bytes (as in ROP) that includes preamble and ID

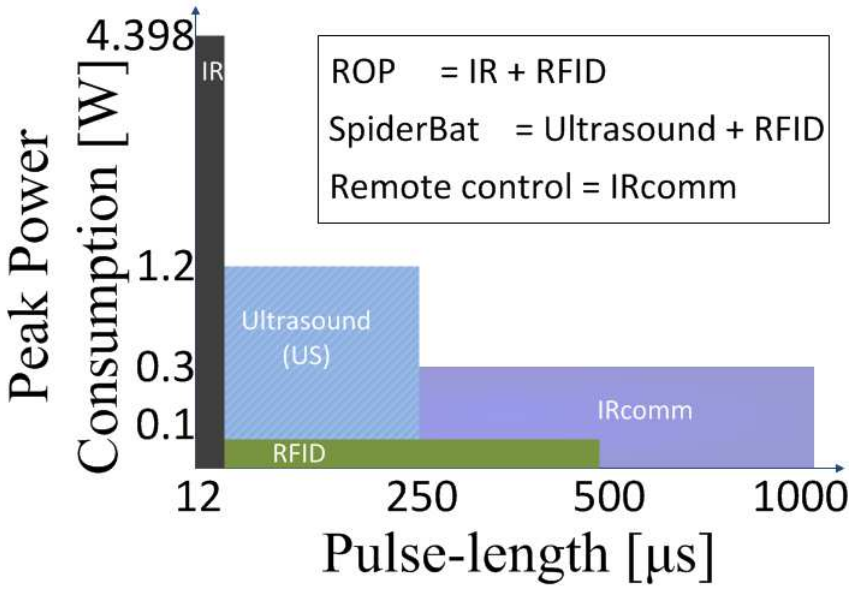
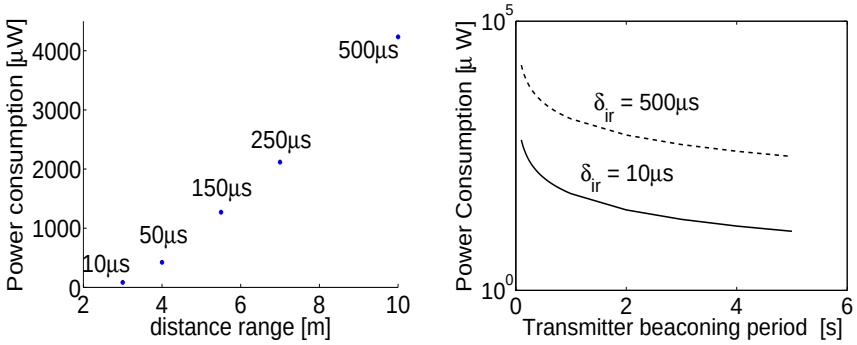


Figure 5.12: Peak transmit power consumption versus pulse-period for ROP, *SpiderBat* [1], and IR remote control technology

(b) Tag Battery Life. Our power measurements indicate that the radio module and IR module consume 32.457μW and 52.78μW respectively. From these measurements, we can a lifetime of 9.854 years when we use the 9V alkaline battery (used in our prototype) of 520mAhrs capacity, and transmit a RF and a 10 μs IR pulse every 1sec. As discussed in Section 5.2.1, this lifetime can be easily extended by a more sophisticated ROP protocol, which we do not consider in this paper.

(c) IR Pulse Length. A large IR pulse length can increase the receiver's view, but it also increases battery power consumption. We plot the measured power consumption versus the corresponding range for different IR pulse duration choices in Figure 5.13(a). We define range as the maximum distance at which the tags can be identified and located through angle and distance estimates by our system. We think, for most augmented reality applications mentioned earlier a 3m range would be sufficient.

(d) Transmitter Beaconing Period. In Figure 5.13 (b) we plot the transmitter power consumption for different beaconing periods τ. The plot indicates a considerable power saving when the beaconing period is increased to 5 seconds but the power savings is less pronounced when a longer IR pulse (500μs) is used (instead of 10μs).



(a) (b)

Figure 5.13: (a) Tag power consumption vs. maximum distance of operation (range of the system) , (b) Tag power consumption vs. beaconing period , for different IR pulse durations.

5.4.3 Receiver Power Consumption

The power consumption at the receiver includes that of the radio plus the IR module. Table 5.3 shows the power consumption of ROP prototype and other existing prototypes. We observe that our prototype receiver performs better than other prototypes. Based on the measurements, the battery life (of a 3V alkaline AA battery) at the receiver is about 2 days. Finally, we note that for ROP, we believe the transmitter battery lifetime is more critical as discussed in Section 5.1. We have been less concerned with optimizing receiver power since a wearable receiver would typically be switched ON only when being used or recharged periodically.

5.5 Discussion

Let us briefly discuss limitations and opportunities for future work.

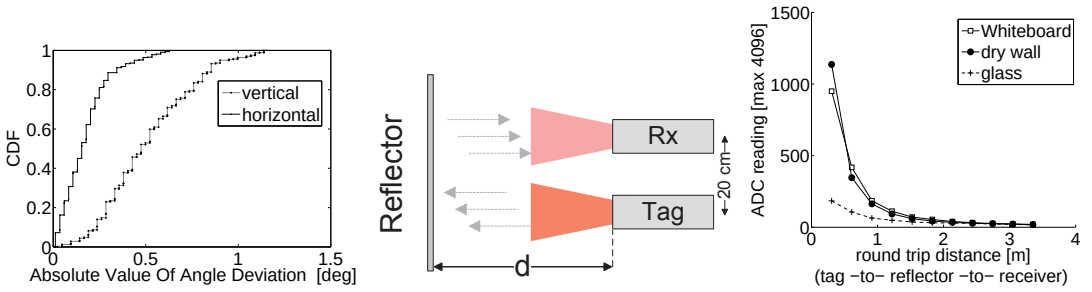
Head movement. For head-mounted wearable receivers, the system only tracks head pose. The object a person is looking at, however, is also affected by eye movement. To understand the effect of head and eye movement, we sought to characterize how consistent head positions are when repeatedly looking at a series of objects. We fitted a laser pointer on the prototype glasses and one of the authors wore this contraption while repeatedly looking at objects on the wall. We recorded a video of the movement of the laser pointer. By analyzing this video footage and knowing the standing position

Technology	**Method**	P_r **[mW]**	**Total** P_r **[mW]**
ROP	IR: AoA (θ, Φ) Radio: ID+sync	9 81	90
RFID	RSSI:proximity and ID (no AoA)	81 (per Rx)	241^{+}
US [1]	US: AoA (θ) Radio: ID+sync	140 100	240
Camera [83]	image:AoA (θ, Φ) ID*	202.2 (1 image)	202.2

Table 5.3: Comparison of receiver average-power consumption (P_r) with other positioning systems [(+ a RSSI based system would require at least 3 receivers for 2D AoA – (θ, Φ), ($*$ uses image recognition, subject to the tagged objects not similar looking, and will require at least two image frames to avoid aliasing)]

and height of the user, we determined the effective angular deviations of the marker from the objects' exact position. We report the cumulative distribution of the data in Figure 5.14 (a). Our experiment indicates a maximum of 1 deg and a median of 0.5 deg angular deviation between the head position and the object the person was looking at. We can infer that a head-mounted system would face this fundamental limit on angular accuracy—any higher precision would require additional eye tracking hardware.

Reflections. Due to the high energy on the IR pulse, reflections from smooth or shiny surfaces, can cause false detection of the beacons on the photodiode receiver. To understand the signal strength of reflected beacons, we let the ROP prototype tag emit 500μs long IR pulses over different distance towards three different reflecting surfaces: whiteboard, glass pane (see through), and dry-wall. The reflections were then received by the photodiodes on the glasses receiver. The setup is illustrated in Figure 5.14 (b). We conducted the experiment with no ambient lighting, to eliminate other potential noise. Figure 5.14 (c) shows the maximum of the three photodiodes' signal strength (as ADC readings) versus the total round-trip distance of the IR signal. Our measurements indicate that the effect of reflections (from typical indoor reflector surfaces) is negligible for round-trip distances greater than 3m. Of course, for a 10μs IR pulse this distance



(a) (b) (c)

Figure 5.14: (a) CDF of angle deviation for head-mounted receivers (experiment repeated for two users of same height), (b). Reflections experiment setup. We used distance of the tag (and Rx) from reflector surface, and accounted for the 20cm spacing, in the round-trip distance computation, (c). IR signal strength from reflections versus round-trip distance

would be much smaller.

Size. The size of our prototype tag, is primarily defined by the size of the RFID tag used, is 3.5cm in the largest dimension. The size of the receiver depends on the placement of the photodiodes in the array, where the spacing and its precision largely affects the accuracy of AoA estimates. We were able to achieve a 1 deg angular accuracy and 9m range with a 3cm spacing between each pair of photodiodes. Our prototype ROP receiver unit is sized ($l \times w \times h$) at 5cm×4cm×3cm. We believe that using surface-mount components on a printed circuit board (PCB) would reduce the size further.

Energy improvements. Reduction in power consumption can be achieved by reducing the pulse length and optimizing the circuit to eliminate noise sources. With this approach it should be possible to achieve larger ranges at about 10μs pulse durations. Detecting optical pulses of 10μs or less duration requires a high speed photo detector and high-speed photo integrator [84]. Here, the mechanical design requires replacing the complex mechanical layout with a carefully designed PCB and appropriate 3D shielding boxes on the board to avoid any electrical pick-ups. The 9V battery in our current design could also be replaced with smaller batteries, such as three coin cells with simple circuit changes.

5.6 Conclusions

In this chapter we presented an approach that help accurate recognition and positioning of objects using light signals. In particular we showed how the visual MIMO approach of array receivers can be used to acquire and track signals from a light transmitter while a radio channel is used for low-power communication. In this work, we designed a hybrid radio-optical beaconing approach that can facilitate accurate and low-power recognition of objects within a user's view, which is particularly useful for many mobile applications including augmented reality. Our approach leverages the high directionality characteristic of an infrared link for precise orientation and distance estimation, and the low power nature of a radio link for synchronization and communication. The novelty of this design lies in the usage of a radio link to synchronize the infrared beacons such that very short high-energy infrared pulses could be used, which results in much reduced energy consumption as well as much simplified receiver design and much smaller receiver size. We prototyped the system by designing radio-optical tags and a wearable receiver, in the form of an object tracking eye-glasses. Our prototype receiver locates the infrared tags with an angular resolution of 1° on the horizontal and vertical dimensions, and up to 9m distance at very low battery power consumption, supporting tag battery life of the order of years. More importantly, the receiver is able to successfully recognize more than 97% of the objects in view. We believe that with such accurate in-view recognition and long lifetime, our system can support a wide range of mobile applications.

Chapter 6

Conclusions

6.1 Summary

In conclusion, this thesis designed a novel communication concept called *visual MIMO* that uses cameras for communication. Specifically, this thesis has made the following contributions:

- **Communication channel model for visual MIMO:** We modeled, analyzed and verified (through experiments) the visual MIMO communications channel and its bounds on information capacity. The model accounts for the unique aspects of the visual channel such as: distortions due to camera perspective, artifacts due to lens (focusing) and motion, quantization and spatial interference from multiple light emitters. Our analysis indicated that visual MIMO, with customized cameras, can allow communication range of hundreds of meters with a relatively wide field-of-view compared to free-space optics, thereby enabling a higher a degree of node mobility. Our analysis of a use-case visual MIMO application of screen-camera communication indicated that such an approach is still promising for medium sized data-transfer or even streaming applications; such as downloading a file from a smartphone screen or streaming a movie from display screens. Our findings indicate that, designing efficient techniques to address perspective distortions is still an open problem for building high-data rate camera communications.

- **Perspective dependent visual MIMO rate adaptation:** With an understanding of the channel impediments on the communication data-rate in visual

MIMO, we designed techniques (for the system) to adapt it's information data-rates towards the detrimental effects of the visual channel distortions. We define a set of operating modes for visual MIMO transmitters and propose a visual MIMO rate adaptation scheme to switch between these modes. Using vehicle-vehicle communication as the motivating application example we proposed three rate-adaptation algorithms for visual MIMO. The algorithms adapt to distance variations by setting a rate corresponding to the best possible spatial pattern of the elements in the transmitter array (*modes*) at that distance. We derive that the *Index VMRA* algorithm that uses a block-CRC indexing scheme to detect occlusion over each light emitting element of the transmitter performs the best of all. This work highlighted the necessity to revisit classical rate adaptation methodologies in RF channels when applied to visual MIMO, due to it unique perspective dependence characteristic.

- **Low-power object recognition using visual MIMO:** In this work we showed how the visual MIMO approach of array receivers can be used to acquire and track signals from a light transmitter while a radio channel is used for low-power communication. This work addressed the object recognition problem by prototyping a battery operated hybrid positioning system that leveraged the visual MIMO idea of array receivers for precise positioning and a side low-power radio control channel for conserving energy. We prototyped the system by designing radio-optical tags and a wearable receiver, in the form of an object tracking eye-glasses. Our prototype receiver locates the infrared tags with an angular resolution of 1° on the horizontal and vertical dimensions, and up to 9m distance at very low battery power consumption, supporting tag battery life of the order of years. More importantly, the receiver is able to successfully recognize more than 97% of the objects in view. We believe that with such accurate in-view recognition and long lifetime, our system can support a wide range of mobile applications.

6.2 End Note

Ongoing advances in science and engineering have largely improved computational capabilities, processing power, reliability, adaptability and usability of communication systems. Motivated by the technological advances, in recent years, in fixed and mobile cyber-physical systems, vehicular technology, and energy management and harvesting, my research through this thesis broadly focused on emerging communications systems in these fields; their theory, modeling, design and application, particular about an emerging technology that will use cameras for communication. This thesis uses the term 'camera'in a very broad sense, as the collection of optical, electrical and mechanical components to sense and interpret light; picture elements or pixels are essentially the light receptors (or photo-receptors) in a camera. Cameras have traditionally been used for capturing images, however, today they have become ubiquitous and pervasively used. Advances in mobile camera technology and processing capabilities, primarily through smartphones, have spurred interest in using cameras for mobile computing through image and video analysis. Driven by the progress and ubiquity of light emitting technology that offer the potential for simultaneous illumination and data-transmission, the recent years have also witnessed an emerging field of wireless communication using visible-light (VLC). With camera applications already becoming pervasive, this thesis envisions a camera to be an integral part of large systems and pervasive applications ranging from mobile and wearable computing to vehicular networks to household and factory robotics in the near future.

References

[1] G. Oberholzer, P. Sommer, and R. Wattenhofer, "Spiderbat: Augmenting wireless sensor networks with distance and angle information," in *ACM/IEEE IPSN*, 2011.

[2] Y. Zang, L. Stibor, H. Reumerman, and H. Chen, "Wireless local danger warning using inter-vehicle communications in highway scenarios," in *Wireless Conference, 2008. EW 2008. 14th European*, 2008, pp. 1–7.

[3] T. ElBatt, S. K. Goel, G. Holland, H. Krishnan, and J. Parikh, "Cooperative collision warning using dedicated short range wireless communications," in *Proceedings of the 3rd international workshop on Vehicular ad hoc networks*. Los Angeles, CA, USA: ACM, 2006, pp. 1–9. [Online]. Available: http://portal.acm.org.proxy.libraries.rutgers.edu/citation.cfm?id=1161066

[4] A. Goldsmith, *Wireless Communications*. Cambridge, 2005.

[5] J. Kahn and J. Barry, "Wireless infrared communications," *Proceedings of the IEEE*, vol. 85, no. 2, pp. 265–298, Feb 1997.

[6] T. Komine and M. Nakagawa, "Fundamental analysis for visible-light communication system using led lights," *IEEE Transactions on Consumer Electronics,*, vol. 50, no. 1, pp. 100–107, Feb 2004.

[7] B. K. P. Horn, *Robot vision*. Cambridge, MA, USA: MIT Press, 1986.

[8] S. Borman, "Raytracing and the camera matrix – a connection," Jun. 2003, a tutorial on the relationships between raytracing formulations of projective geometry and the standard camera matrix representation.

[9] A. Tang, J. Kahn, and K.-P. Ho, "Wireless infrared communication links using multi-beam transmitters and imaging receivers," in *Communications, 1996. ICC 96, Conference Record, Converging Technologies for Tomorrow's Applications. 1996 IEEE International Conference on*, vol. 1, Jun 1996, pp. 180–186 vol.1.

[10] G. J. Foschini and M. J. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Comunications: Kluwer Academic*, vol. 6, no. 3, pp. 311–355, Mar. 1998.

[11] "Mipav," http://mipav.cit.nih.gov/documentation/HTML%20Algorithms/FiltersSpatialGaussianBlur.html.

[12] "Stan moore astronomy," http://www.stanmooreastro.com/pixel_size.html.

[13] A.Ashok, M.Gruteser, N. B. Mandayam, J. Silva, K. Dana, and M.Varga, "Challenge: Mobile optical networks through visual mimo," in *MobiCom '10: Proceedings of the sixteenth annual international conference on Mobile computing and networking*. New York, NY, USA: ACM, 2010, pp. 105–112.

[14] L. Zheng and D. N. C. Tse, "Diversity and multiplexing: A fundamental tradeoff in multiple-antenna channels," *IEEE Trans. Inform. Theory*, vol. 49, no. 5, pp. 1073–1096, May 2003.

[15] D. N. C. Tse, P. Vishwanath, and L. Zheng, "Diversity-multiplexing tradeoff in multiple-access channels," *IEEE Trans. Inform. Theory*, vol. 50, no. 9, pp. 1859–1874, Sep. 2004.

[16] B. Nakhkoob, M. Bilgi, M. Yuksel, and M. Hella, "Multi-transceiver optical wireless spherical structures for manets," *IEEE J.Sel. A. Commun.*, vol. 27, no. 9, pp. 1612–1622, 2009.

[17] T. Saito, S. Haruyama, and M. Nakagawa, "Inter-vehicle communication and ranging method using led rear lights," *Proceedings of the fifth IASTED international conference on Communication Systems and Networks*, vol. 5, pp. 278–283, Aug 2006.

[18] B. Mukherjee, *Optical WDM Networks*. Springer, Jan 2006.

[19] S. Muhammad, P. Kohldorfer, and E. Leitgeb, "Channel modeling for terrestrial free space optical links," in *Transparent Optical Networks, 2005, Proceedings of 2005 7th International Conference*, vol. 1, July 2005, pp. 407–410 Vol. 1.

[20] J. c. Arun Majumdar, *Free-Space Laser Communications:Principles and Advances*. Springer, Nov 2007.

[21] S. S. D. Neo, "Free space optics communication for mobile military platforms," Master?s thesis, Naval Postgraduate School, Tech. Rep., 2003.

[22] M. Yuksel, J. Akella, S. Kalyanaraman, and P. Dutta, "Free-space-optical mobile ad hoc networks: Auto-configurable building blocks," *Wirel. Netw. 15*, vol. 15, no. 3, pp. 295–312, 2007.

[23] E. Fischer, P. Adolph, T. Weigel, C. Haupt, and G. Baister, "Advanced optical solutions for inter-satellite communications," *Optik - International Journal for Light and Electron Optics*, vol. 112, no. 9, pp. 442 – 448, 2001. [Online]. Available: http://www.sciencedirect.com/science/article/B7GVT-4DPXHWJ-2S/2/fbb7f172a9fcc25cce1fa07823c41380

[24] W. Ng, A. Walson, G. Tangonan, J. Lee, and I. Newberg, "Optical steering of dual band microwave phased array antenna using semiconductor laser switching," *Electronics Letters, IEEE-*, vol. 26, no. 12, pp. 791–793, June 1990.

[25] H. Willebrand and B. Ghuman, *Free Space Optics:Enabling Optical Connectivity in Today's Networks*. Sams, 2002.

[26] "Ubiquitous communication laboratory," http://haruyama.sdm.keio.ac.jp/ubiquitous/index.html.

[27] G. Pang, T. Kwan, H. Liu, and C.-H. Chan, "Led wireless," *Industry Applications Magazine, IEEE*, vol. 8, no. 1, pp. 21–28, Jan/Feb 2002.

[28] "Visible light communication consortium," http://vlcc.net.

[29] M. S. Navin Kumar, Nuno Lourenco and R. Aguiar, "Visible light communication systems conception and vidas," *IETE Tech. Review*, vol. 25, pp. 359–367, 2008.

[30] S. Kitano, S. Haruyama, and M. Nakagawa, "Led road illumination communications system," in *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 5, Oct. 2003, pp. 3346–3350.

[31] G. Pang, K.-L. Ho, T. Kwan, and E. Yang, "Visible light communication for audio systems," *Consumer Electronics, IEEE Transactions on*, vol. 45, no. 4, pp. 1112–1118, Nov 1999.

[32] "Smart lighting erc at boston university (slc/bu)," http://smartlighting.bu.edu/research/index.html.

[33] J. Carruthers, S. Carroll, and P. Kannan, "Propagation modelling for indoor optical wireless communications using fast multi-receiver channel estimation," *Optoelectronics, IEE Proceedings* -, vol. 150, no. 5, pp. 473–481, Oct. 2003.

[34] "Exploding interest in visible light communications: An applications viewpoint," http://www.bu.edu/systems/files/2009/07/Little-Smartlight-compressed.pdf.

[35] H. Sugiyama, S. Haruyama, and M.Nakagawa, "Experimental investigation of modulation methods for visible light communications," *IEEE Transactions on Communications,*, vol. 89, no. 12, pp. 3393–3400, Dec 2006.

[36] H. Binti Che Wook, T. Komine, S. Haruyama, and M. Nakagawa, "Visible light communication with led-based traffic lights using 2-dimensional image sensor," in *Consumer Communications and Networking Conference, 2006. CCNC 2006. 3rd IEEE*, vol. 1, Jan. 2006, pp. 243–247.

[37] S. Arai, S. Mase, T. Yamazato, T. Endo, T. Fujii, M. Tanimoto, K. Kidono, Y. Kimura, and Y. Ninomiya, "Experimental on hierarchical transmission scheme for visible light communication using led traffic light and high-speed camera," in *Vehicular Technology Conference, 2007. VTC-2007 Fall. 2007 IEEE 66th*, 30 2007-Oct. 3 2007, pp. 2174–2178.

[38] A. Mohan, G. Woo, S. Hiura, Q. Smithwick, and R. Raskar, "Bokode: imperceptible visual tags for camera based interaction from a distance," in *SIGGRAPH '09: ACM SIGGRAPH 2009 papers*. New York, NY, USA: ACM, 2009, pp. 1–8.

[39] S. Jivkova, B. Hristov, and M. Kavehrad, "Power-efficient multispot-diffuse multiple-input-multiple-output approach to broad-band optical wireless communications," *Vehicular Technology, IEEE Transactions on*, vol. 53, no. 3, pp. 882 – 889, may. 2004.

[40] L. Zeng, D. C. O'Brien, H. Minh, G. E. Faulkner, K. Lee, D. Jung, Y. Oh, and E. T. Won, "High data rate multiple input multiple output (mimo) optical wireless communications using white led lighting," *IEEE J.Sel. A. Commun.*, vol. 27, no. 9, pp. 1654–1662, 2009.

[41] S. Hranilovic and F. Kschischang, "A pixelated mimo wireless optical communication system," *Selected Topics in Quantum Electronics, IEEE Journal of*, vol. 12, no. 4, pp. 859 –874, jul. 2006.

[42] C. Liang, M. Garfield, and K. R. D. T. P. Kurzweg, "Mimo space-time coding for diffuse optical communication," *Microwave and Optical Technology Letters*, vol. 48, pp. 1108 – 1110, may. 2006.

[43] H. Lomheim, "Cmos/ccd sensors and camera systems."

[44] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[45] "Camera calibration toolbox for matlab," http://www.vision.caltech.edu/bouguetj/calib_doc/index.html#links.

[46] S. D. Perli, N. Ahmed, and D. Katabi, "Pixnet: interference-free wireless links using lcd-camera pairs," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, ser. MobiCom '10. New York, NY, USA: ACM, 2010, pp. 137–148. [Online]. Available: http://doi.acm.org/10.1145/1859995.1860012

[47] T. Hao, R. Zhou, and G. Xing, "Cobra: color barcode streaming for smartphone systems," in *Proceedings of the 10th international conference on Mobile systems, applications, and services*, ser. MobiSys '12. New York, NY, USA: ACM, 2012, pp. 85–98. [Online]. Available: http://doi.acm.org/10.1145/2307636.2307645

[48] "Billboard sizes," http://www.sbuilts.com/sizes.cfm.

[49] H. Chen, R. Sukhthankar, G. Wallace, and T. jen Cham, "Calibrating scalable multi-projector displays using camera homography trees," in *In Computer Vision and Pattern Recognition*, 2001, pp. 9–14.

[50] R. Yang, D. Gotz, J. Hensley, H. Towles, and M. S. Brown, "Pixelflex: A reconfigurable multi-projector display system," 2001.

[51] X. Liu, D. Doermann, and H. Li, "A camera-based mobile data channel: capacity and analysis," in *Proceedings of the 16th ACM international conference on Multimedia*, ser. MM '08. New York, NY, USA: ACM, 2008, pp. 359–368. [Online]. Available: http://doi.acm.org/10.1145/1459359.1459408

[52] "High capacity color barcodes," http://research.microsoft.com/en-us/projects/hccb/about.aspx.

[53] S. H. Y. Wong, H. Yang, S. Lu, and V. Bharghavan, "Robust rate adaptation for 802.11 wireless networks," in *in ACM Mobicom*, 2006, pp. 146–157.

[54] K. Ramach, H. Kremo, M. Gruteser, and P. Spasojevi, "Scalability analysis of rate adaptation techniques in congested ieee 802.11 networks: An orbit testbed comparative study," in *in Proc. of IEEE WoWMoM*, 2007.

[55] B. Sadeghi, V. Kanodia, A. Sabharwal, and E. Knightly, "Oar: an opportunistic auto-rate media access protocol for ad hoc networks," *Wirel. Netw.*, vol. 11, pp. 39–53, January 2005. [Online]. Available: http://dx.doi.org/10.1007/s11276-004-4745-x

[56] R. T. Morris, J. C. Bicket, and J. C. Bicket, "Bit-rate selection in wireless networks," Masters thesis, MIT, Tech. Rep., 2005.

[57] S. Nanda, K. Balachandran, and S. Kumar, "Adaptation techniques in wireless packet data services," *Communications Magazine, IEEE*, vol. 38, no. 1, pp. 54 –64, Jan. 2000.

[58] I. Pefkianakis, Y. Hu, S. H. Wong, H. Yang, and S. Lu, "Mimo rate adaptation in 802.11n wireless networks," in *Proceedings of the sixteenth annual international conference on Mobile computing and networking*, ser. MobiCom '10. New York, NY, USA: ACM, 2010, pp. 257–268. [Online]. Available: http://doi.acm.org/10.1145/1859995.1860025

[59] Q. Xia, M. Hamdi, and K. Ben Letaief, "Open-loop link adaptation for next-generation ieee 802.11n wireless networks," *Vehicular Technology, IEEE Transactions on*, vol. 58, no. 7, pp. 3713 –3725, 2009.

[60] L. Diana and J. Kahn, "Rate-adaptive modulation techniques for infrared wireless communications," in *Communications, 1999. ICC '99. 1999 IEEE International Conference on*, 1999.

[61] A. García-Zambrana, C. Castillo-Vázquez, and B. Castillo-Vázquez, "Rate-adaptive fso links over atmospheric turbulence channels by jointly using repetition coding and silence periods," *Opt. Express*, vol. 18, no. 24, pp. 25 422–25 440, Nov 2010. [Online]. Available: http://www.opticsexpress.org/abstract.cfm?URI=oe-18-24-25422

[62] J. Grubor, V. Jungnickel, and K.-D. Langer, "Rate-adaptive multiple sub-carrier-based transmission for broadband infrared wireless communication," in *Optical Fiber Communication Conference, 2006 and the 2006 National Fiber Optic Engineers Conference. OFC 2006*, 2006, p. 10 pp.

[63] "Wikitude : The world's leading augmented reality sdk," http://www.wikitude.com/, 2013. [Online]. Available: http://www.wikitude.com/

[64] H. Wang, S. Sen, A. Elgohary, M. Farid, M. Youssef, and R. R. Choudhury, "No need to war-drive: unsupervised indoor localization," in *ACM MobiSys*, 2012.

[65] "Inside Microsoft Research: making purchases with zero effort," http://tinyurl.com/bkjt9js, 2013. [Online]. Available: http://tinyurl.com/bkjt9js

[66] G. B. Airy, "On the diffraction of an object-glass with circular aperture," *Transactions of the Cambridge Philosophical Society*, 1835.

[67] R. Want, A. Hopper, V. Falcão, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems*, 1992.

[68] "Bytelight : Indoor location. with light," http://www.bytelight.com/. [Online]. Available: http://www.bytelight.com/

[69] "Wiimote," http://wiibrew.org/wiki/Wiimote. [Online]. Available: http://wiibrew.org/wiki/Wiimote

[70] "Irda," http://http://www.irda.org/index.cfm/.

[71] "Giga-ir high speed opto-communication," http://www.google.com/glass/start/. [Online]. Available: http://www.embednet.com/Giga-IR_General.pdf

[72] E. Shih, P. Bahl, and M. J. Sinclair, "Wake on wireless: an event driven energy saving strategy for battery operated devices," in *ACM MobiCom*, 2002.

[73] G. Ananthanarayanan and I. Stoica, "Blue-fi: enhancing wi-fi performance using bluetooth signals," in *ACM MobiSys*, 2009.

[74] M. Uddin and T. Nadeem, "A2psm: audio assisted wi-fi power saving mechanism for smart devices," in *ACM HotMobile*, 2013.

[75] C. Yuodelis and A. Hendrickson, "A qualitative and quantitative analysis of the human fovea during development," *Vision Research*, 1986.

[76] K. Whitehouse, C. Karlof, and D. Culler, "A practical evaluation of radio signal strength for ranging-based localization," *ACM SIGMOBILE MCCR*, 2007.

[77] D. Zhang, F. Xia, Z. Yang, L. Yao, and W. Zhao, "Localization technologies for indoor human tracking," in *IEEE FutureTech*, 2010.

[78] P. Bahl and V. Padmanabhan, "Radar: an in-building rf-based user location and tracking system," in *IEEE INFOCOM*, 2000.

[79] L. Korba, S. Elgazzar, and T. Welch, "Active infrared sensors for mobile robots," *IEEE Transactions on Instrumentation and Measurement*, 1994.

[80] G. Benet, F. Blanes, J. E. Simó, and P. Pérez, "Map building using infrared sensors in mobile robots," *Book chapter in: New Developments in Robotics Research? Nova Science Publishers, Inc*, 2006.

[81] "Tsal 5300 near-ir led," www.vishay.com/docs/81008/tsal5300.pdf. [Online]. Available: http://www.vishay.com/docs/81008/tsal5300.pdf

[82] "Bp10nf si photodiode," http://www.vishay.com/docs/81503/bp10nf.pdf. [Online]. Available: http://www.vishay.com/docs/81503/bp10nf.pdf

[83] R. LiKamWa, B. Priyantha, M. Philipose, L. Zhong, and P. Bahl, "Energy characterization and optimization of image sensing toward continuous mobile vision," in *ACM MobiSys*, 2013.

[84] J. Williams, 1991, linear Technologies:High Speed Amplifier Techniques, A Designers Companion for Wideband Circuitry.

Appendix A

List of Refereed Publications as a Ph.D. student

(C : conference, D : demo, J : journal, in chronological order)

C-1 *Capacity of Pervasive Camera Based Communication Under Perspective Distortion*, Ashwin Ashok, Shubham Jain, Marco Gruteser, Narayan Mandayam, Wenjia Yuan, Kristin Dana, **IEEE PerCom 2014**

C-2 *Phase Messaging Method for Time-of-flight Cameras*, Wenjia Yuan, Kristin Dana, Ashwin Ashok, Rich Howard, Ramesh Raskar, Marco Gruteser, and Narayan Mandayam, **IEEE International Conference on Computational Photography, ICCP 2014**

D-3 *Demo: BiFocus – Using Radio-Optical Beacons for An Augmented Reality Search Application*, Ashwin Ashok, Chenren Xu, Tam Vu, Marco Gruteser, Richard Howard, Yanyong Zhang, Narayan Mandayam, Wenjia Yuan, Kristin Dana, **ACM MobiSys 2013**

C-4 *Spatially Varying Radiometric Calibration For Camera-Display Messaging*, Wenjia Yuan, Kristin Dana, Ashwin Ashok, Marco Gruteser, Narayan Mandayam, **IEEE Global Conference on Signal and Image Processing (GlobalSIP) Symposium on Mobile Imaging, Dec 2013**

C-5 *Photometric Modeling for Active Scenes*, Wenjia Yuan, Kristin Dana, Ashwin Ashok, Marco Gruteser, Narayan Mandayam, **IEEE CVPR Workshop on Computational Cameras and Displays, Poster Presentation, 2013**

D-6 *Demo: User Identification and Authentication with Capacitive Touch Communication*, Tam Vu, Ashwin Ashok, Akash Baid, Marco Gruteser, Jeffrey Walling,

Predrag Spasojevic, Richard Howard, **ACM MobiSys 2012**

C-7 *Dynamic and Invisible Messaging for Visual MIMO*, Wenjia Yuan, Kristin Dana, Ashwin Ashok, Marco Gruteser, Narayan Mandayam, **IEEE Workshop on Applications In Computer Vision (WACV) 2012**

C-8 *Computer Vision Methods for Visual MIMO Optical System*, Wenjia Yuan, Kristin Dana,Ashwin Ashok, Marco Gruteser, Narayan Mandayam, Michael Varga, **IEEE Computer Vision and Pattern Recognition (CVPR) Workshop 2011**

D-9 *Demo: Visual MIMO based LED-camera communication Applied to Automobile Safety*, Michael Varga, Ashwin Ashok (co-primary), Marco Gruteser, Narayan Mandayam, Wenjia Yuan, Kristin Dana, **ACM MobiSys 2011**

C-10 *Rate Adaptation in Visual MIMO*, Ashwin Ashok, Marco Gruteser, Narayan Mandayam, Ted Kwon, Wenjia Yuan, Kristin Dana, **IEEE SECON 2011**

C-11 *Characterizing Multiplexing and Diversity in Visual MIMO*, Ashwin Ashok, Marco Gruteser, Narayan Mandayam, Kristin Dana, **IEEE CISS 2011**

C-12 *Challenge: Mobile Optical Networks Through Visual MIMO*, Ashwin Ashok, Marco Gruteser, Narayan Mandayam, Jayant Silva, Michael Varga, Kristin Dana, **ACM MobiCom 2010**