Pixnet: Designing Interference-free Wireless Links using LCD-Camera Pairs

by

Samuel David Perli

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Master of Science

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2010

Author .....
Department of Electrical Engineering and Computer Science
May 21, 2010

Certified by .....
Dina Katabi
Associate Professor of Computer Science
Thesis Supervisor

Accepted by .....
Terry Orlando
Chairman, Department Committee on Graduate Students

Pixnet: Designing Interference-free Wireless Links using LCD-Camera Pairs

by

Samuel David Perli

Submitted to the Department of Electrical Engineering and Computer Science
on May 21, 2010, in partial fulfillment of the
requirements for the degree of
Master of Science

Abstract

Given the abundance of cameras and LCDs in today's environment, there exists an untapped opportunity for using these devices for communication. Specifically, cameras can tune to nearby LCDs and use them for network access. The key feature of these LCD-camera links is that they are highly directional and hence enable a form of interference-free wireless communication. This makes them an attractive technology for dense, high contention scenarios. The main challenge however, to enable such LCD-camera links is to maximize coverage, that is to deliver multiple Mb/s over multi-meter distances, independent of the view angle. To do so, these links need to address unique types of channel distortions such as perspective distortion and blur.

This thesis explores this novel communication medium and presents PixNet, a system for transmitting information over LCD-camera links. PixNet generalizes the popular OFDM transmission algorithms to address the unique characteristics of the LCD-Camera links which include perspective distortion, blur and sensitivity to ambient light. We built a prototype of PixNet using off-the-shelf LCDs and cameras. An extensive evaluation shows that a single PixNet link delivers data rates of up to 12 Mb/s at a distance of 10 meters, and works with view angles as wide as 120°.

Thesis Supervisor: Dina Katabi
Title: Associate Professor of Computer Science

Acknowledgments

I give thanks to the Lord, my God, for granting me wisdom, spirit of understanding, strength, faith and perseverance to shepherd me through this journey of life. I thank my parents for their exemplary lives, deep love, prayers and support. My father, especially for his encouragement and counsel. My mother, for her loving care and uplifting spirit. I thank my advisor, prof. Dina Katabi, for her insights and guidance. It has been a wonderful experience working along with her. I am grateful to my labmate Nabeel for the tremendous help and support he has been to me. It was fun working and playing along with him. I thank my small group, for their heartfelt fellowship, prayers and awesome potlucks. I am indebted to my undergrad mentor, Prof. David Koilpillai, for the indelible impression he left on me. Finally, I thank my brother, my sister and all my friends near and far, who constantly remember me in their prayers.

Contents

List of Figures

List of Tables

Chapter 1

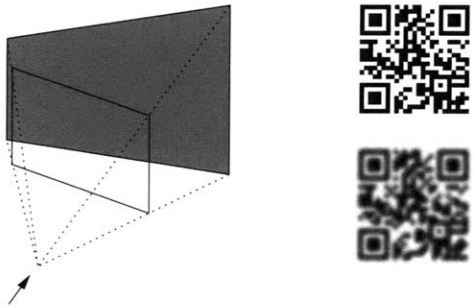
Introduction

Cameras and LCDs are abundant in today's environment, both in stand-alone form and integrated form - embedded in laptops, smart phones, and PDAs. This abundance creates an untapped opportunity for using these devices for wireless communication. For example, LCDs mounted on walls or ceilings can encode data into visual frames, allowing camera-equipped devices to download this information. The key feature of such LCD-camera links is that they are interference-free. This is due to the short wavelengths in the visible light spectrum that makes the communication link highly directional. Thus, a multitude of such links can operate simultaneously in a dense area, such as in a conference scenario or a hotspot. Such LCD-camera links can potentially evolve into a new wireless technology that is useful in dense high-contention scenarios, similar to how Bluetooth targets low-power scenarios, and whitespaces target long-range communication.

1.1 Challenges

While they offer new opportunities, LCD-camera links bring about new challenges. Specifically, an LCD-camera link exhibits three main types of distortions:

- *Perspective distortion:* Since they operate in the visible light spectrum, LCD-camera links require line of sight. This requires the designer to pay special attention to where the LCDs are placed in order to maximize coverage. If an LCD and a camera can communicate only when they are perfectly aligned, the resulting system is less likely to be broadly applicable to different usage scenarios. In contrast, if an LCD and a camera can communicate in the presence of view angles, similar to how a human sees



(a) Perspective Distortion

(b) Blur

Figure 1-1—Example distortions of the LCD-Camera channel. Perspective distortion occurs when the camera and the LCD are at an angle relative to each other. Blur occurs as a result of lack-of-focus or movement while capturing an image.

a screen even when he looks at it from an angle, coverage is significantly extended. The challenge is that the image of a rectangular screen becomes a trapezoid when viewed from an angle, as shown in Fig. 1-1(a). As a result, some pixels on the LCD screen expand at the camera, while others shrink. The more coverage one would like from an LCD-camera link, the greater perspective distortion these links must be able to tolerate.

- *Blur:* Any handshaking or movement while capturing an image or a lack of focus can introduce blur in the image, which causes the pixels to blend together, as in Fig. 1-1(b). An LCD-camera communication system must be able to deal with such blending and still successfully recover the transmitted bits.
- *Ambient Light:* Ambient light is a source of noise for LCD-camera links because it changes the luminance of the received pixels. This can cause errors in the information encoded in the pixels, resulting in information loss at the receiver.

Thus, the LCD-camera channel needs a new transmission scheme that can handle the above distortions, that are significantly different from the distortions seen in RF wireless channels.

Past work in the area of computer graphics has looked at these problems in the context of 2D barcodes, e.g., QR code [2] or Data matrix [3]. These codes are printed on walls or objects. Users with a camera phone can take a picture of these barcodes, decode them, and obtain a description of the attached object or surrounding space [25, 26]. Barcodes however have relatively low information density and must be read at close proximity [12, 19]. In contrast, we aim to develop an LCD-camera link that supports high data rates at multi-meter distances and wide view angles.

1.2 Key features of our approach

This thesis presents PixNet, a system for transmitting information over LCD-camera links. In contrast to all past work on 2D barcodes, which encode information directly in the visual domain, PixNet encodes information in the *frequency domain*. Such a design is inspired by the popular OFDM transmission scheme, widely used in modern RF technologies. However, unlike existing RF-based OFDM schemes that encode data in time frequencies, PixNet encodes data in two-dimensional spatial frequencies. More importantly, PixNet generalizes OFDM receiver algorithms to deal with the unique distortions of the LCD-camera link. Using PixNet we show that such a generalized frequency-based design provides a unified framework to deal with the distortions in the LCD-camera channel.

PixNet has the following three components:

(a) Perspective Correction Algorithm: A picture taken by a digital camera is a sampled version of the captured object. Perspective distortion occurs when the sampling frequency is irregular. For example, a rectangular screen becomes a trapezoid if the columns on the right are sampled at a lower frequency (i.e., with more pixels) than those on the left (Fig. 1-1(a)). Since PixNet operates in the frequency domain, it naturally addresses irregularity in the sampling frequencies. Specifically, PixNet generalizes the OFDM algorithm for correcting the sampling frequency offset (SFO) between sender and receiver to allow it to work with irregular sampling offsets. Once the receiver knows the sampling frequency offset in each part of the image, it re-samples the image at the right frequencies to correctly recover the bits encoded in the frame.

(b) Blur-Adaptive Coding: Approaches that encode bits directly in the visual domain, like 2D barcodes, fail quickly in the presence of blur because the bits blend together. In contrast, since PixNet encodes information in the frequency domain, it is more resilient to blur. Blur, in the frequency domain, translates into attenuation in the high frequencies while the low frequencies remain intact. Therefore, PixNet naturally identifies the frequencies affected by blur and prevents the error from spreading into other bits. PixNet treats different frequencies differently: frequencies that are badly attenuated due to blur are not used for transmitting information; frequencies mildly affected by blur are used for transmitting information but protected with a high redundancy error correction code; frequencies not affected by blur are used for transmission and protected with a low redundancy error correcting code.

(c) Ambient Light Filter: Approaches that encode information directly in the visual domain have to

perform a special preprocessing step referred to as light balancing [14]. In contrast, PixNet operates in the frequency domain. Since ambient light changes the overall luminance, it only affects the DC frequency. Thus, PixNet can filter out the impact of ambient light simply by ignoring the DC frequency.

1.3 Result highlights

We built a software prototype of PixNet and evaluated it using commodity LCDs and cameras. Empirical results using Dell 30 inch screens with consumer cameras such as Casio EX-F1 and Nikon D3X reveal the following findings:

- Using PixNet a single LCD-camera link can deliver data rates of up to 12 Mb/s, at a distance of 10 meters.
- PixNet's links support wide view angles. Specifically, PixNet delivers 8 Mb/s at view angles as wide as 120°.
- We also compare PixNet to a baseline system that uses the popular Quick Response (QR) 2D barcode and stacks as many of them as can fit in an LCD frame. Our results show that in comparison with QR codes, PixNet delivers up to $2x - 9x$ higher throughput depending on the distance and can also tolerate $3x$ wider view angles.

1.4 Contributions

This thesis makes the following contributions:

- To the best of our knowledge, PixNet is the first system where LCD-camera links are shown in a working deployment to deliver high throughput data rates over multi-meter distances and wide view angles.
- PixNet presents a novel OFDM-based receiver algorithm that addresses perspective distortion.
- PixNet presents a blur-adaptive error correction code that is robust to distortions in the LCD-camera channel.
- We present an extensive empirical study of data communication over LCD-camera links that addresses distances, view angles, camera focus, and different approaches for encoding information, e.g., 2D OFDM and 2D barcodes.

Chapter 2

Related Work

Coding information visually has primarily been in the form of barcodes, which are visual markers that are printed on commercial goods for stock control or painted on walls or roads to convey location information. Barcodes are typically decoded using flying spot scanning lasers and a single photodetector. This minimizes perspective distortion and blur [23]. Recent work on pervasive computing has developed a new class of 2D barcodes that are captured and processed using cell phone cameras. A few such barcodes are shown in Fig. 2-1 [2, 3, 4, 1]. Today there is a spectrum of 2D barcodes. Some 2D barcodes have been proposed for information storage [10]. Though they advertise high information density, due to their proprietary nature, little is known about their tolerance to blur, perspective distortion and distances. Other 2D codes (e.g., QR codes, Data Matrix, and BoKodes) are shown to be robust to visual distortion and blur but are limited in terms of their information density [17]. There has also been some work on time-multiplexing such 2D barcodes to allow decoding across multiple frames. However, these schemes deliver only a few hundred characters per minute and do not deal with large distances or wide view angles. PixNet's design differs from all of these codes as it encodes information in the frequency domain using spatial OFDM. As a result, it achieves higher data rates and wider view angles as shown in §9.

PixNet builds on a proposal by Hranilovic and Kschischang [11] that also advocate using OFDM to transmit between an LCD-camera pair. Their approach however uses OFDM as-is and does not deal with perspective distortion, blur, or frame synchronization. As a result, it applies only to scenarios in which the camera and LCD face each other and are perfectly aligned and are in perfect focus. Additionally, the work in [11] is based on emulation, i.e., the channel function is measured empirically but no data is physically transmitted. The channel is then applied to the OFDM encoded data in simulation. In contrast, PixNet

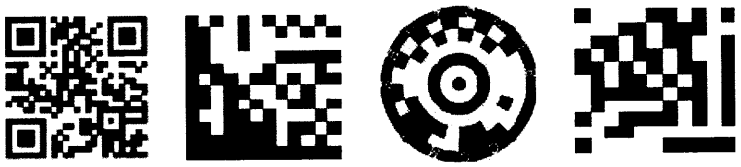


Figure 2-1—A few examples of 2D barcodes used in mobile tagging. From left: QR Codes, Data Matrix Codes, Shot Codes and EZ Codes.

generalizes OFDM's sampling correction algorithm to address perspective distortion and augments it with a blur-adaptive error correction code. Moreover, PixNet extends its design to three dimensions as well. PixNet also presents the first implementation of an LCD-camera communication system that delivers high data rates over multi-meter distances and wide view angles. Our work is also related to [27] in which digital information is encoded directly in the visual domain. PixNet differs from [27] as it encodes information in the frequency domain by generalizing OFDM.

PixNet is also related to work on free-space optics (FSO) and visible light communication (VLC). Free space optics (FSO) use laser beams that transmit light in the atmosphere without the use of physical fiber. This technology was originally developed by the military but has since been used in optical interconnects for on-board and cross-board communication [9, 15]. FSO, however, is not applicable in the end user scenarios we are targeting, because it requires almost perfect alignment of the transmitter and the receiver.

For indoor environments, visible light communication (VLC) has been proposed [18], where light sources (e.g., light bulbs) are modulated to enable data communication in addition to illumination [13, 16]. Such light sources are highly diffusive in nature, resulting in very low throughput communication. Although there has been some work on specialized sources and photodetectors [16] that address this problem, VLC is fundamentally not appropriate for the high bandwidth scenarios targeted by PixNet.

Chapter 3

An Overview of PixNet

PixNet is a communication system that involves design techniques at both the transmitter and the receiver. It involves modulation and coding techniques that map raw digital bit stream on to pixel intensity values, algorithms that achieve accurate sampling at the receiver while correcting for various distortions, demodulation and decoding techniques that convert pixel intensity values back in to a digital bit stream.

To put the scope of this work into perspective, we start by listing the assumptions under which we have developed PixNet.

- We consider scenarios where the camera is already setup to take pictures of the communicating LCD screen. In static environments the user may focus the camera on the screen. In dynamic scenarios, one can leverage prior work on steerable projector-camera systems [8, 22] and use automatic focusing. Note that this does not mean that the LCD and camera are aligned; the camera can still be viewing the LCD from any view angle.
- We focus on one-way communication from an LCD to a camera. Higher layer protocols and applications may require two-way communication (e.g., TCP Acks). A full-featured integration of LCD-camera links into a general purpose networking framework has to consider these issues in more detail. In this thesis however, we assume that Acks and uplink traffic are sent over the RF WLAN and focus only on one-way transmission.

Fig. 3-1 shows various blocks that make up PixNet's transmit chain. As in traditional communication systems, the binary data to be transmitted is first encoded for error protection. PixNet uses Reed Solomon error correction codes that introduce a redundancy of $2x$ bits in order to tolerate x bit errors. However, as

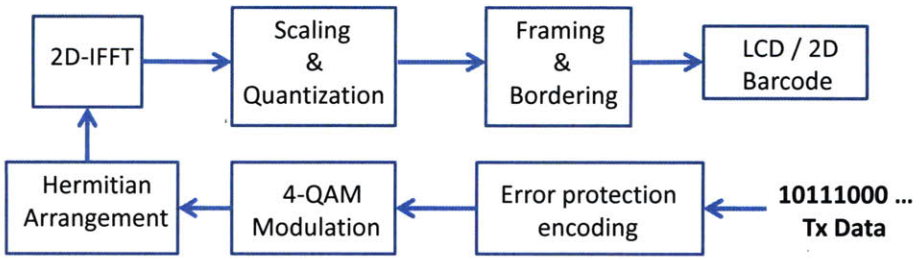


Figure 3-1—PixNet's Transmit chain.

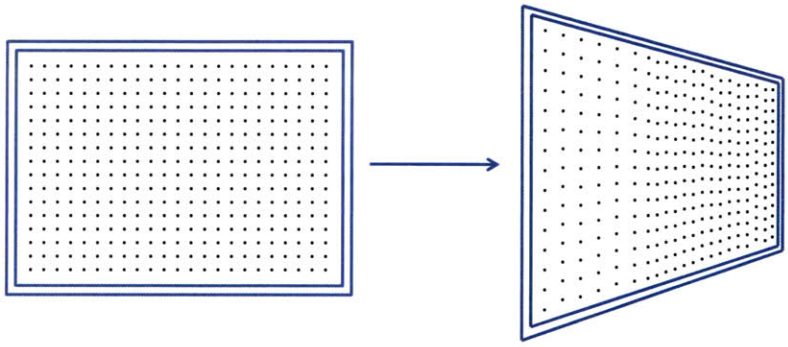
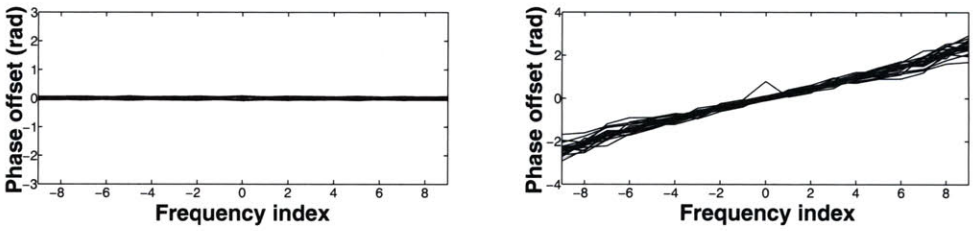


Figure 3-2—An example illustrating perspective distortion. In the presence of distortion, the receiver samples the transmitted image in an irregular fashion.



(a) Phase offsets in the absence of distortion

(b) Phase offsets in the presence of distortion

Figure 3-3—Calculated phase offsets at the receiver. These phase offsets are used in a feed-back loop for finding the optimal sampling.

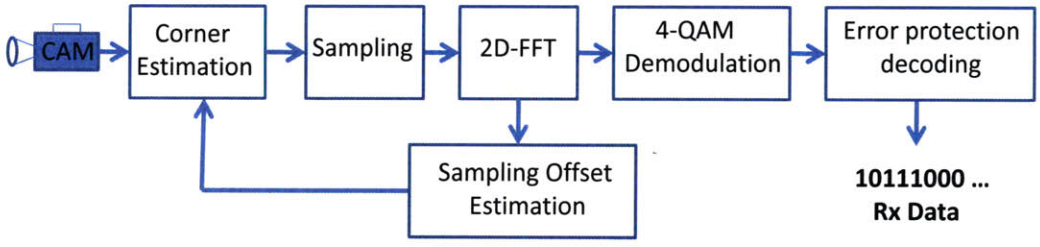


Figure 3-4—PixNet's Receive chain.

will be described in §6, different bits are encoded with different amounts of tolerence to bit errors. Since PixNet transmits data in the frequency domain similar to an OFDM system, it performs a 2D-IFFT of the data to be transmitted. But the output of an IFFT is in general complex and cannot be represented in terms of pixel intensity values which are real. To deal with this issue, PixNet arranges the modulated data in a Hermitian format. According to the properties of the Fourier Transform, if the input to an IFFT is Hermitian, the output is real. The output of the IFFT is then suitably scaled and quantized to match the 8-bit resolution of current displays. After adding cyclic prefix and a frame border, the data is then displayed on the LCD.

One might imagine that once an image of the displayed frame is captured, frame boundaries and/or visual markers could be used to identify the corners of the frame. The frame could then be sampled appropriately to retrieve the transmitted pixels. This approach doesn't work because of the following reasons:

- Depending on the location of the camera relative to the LCD, the imaged frame undergoes a perspective distortion as shown in Fig. 3-2. The received image is no longer rectangular and the transmitted pixels map to irregular locations in the received image.
- Due to inter-pixel interference, the image experiences blur and as a result, there is no one-to-one mapping between the transmitted pixels and the received pixels.

The above challenges make it difficult to find the optimal sampling instants in the received image. But since PixNet operates in the frequency domain, this problem is easily addressed. After detecting the four corners of the frame, the optimal sampling instants can be found by using the properties of Fourier Transform. As will be shown in §5, any irregularity in the sampling instants at the receiver result in linearly varying phase-offsets between the transmitted and the received values. Fig. 3-3 shows the phase offsets for the case when the image is optimally sampled and for the case when the image is non-optimally sampled. As described in §5, these phase offsets are calculated at the receiver using *pilot* bins. If the calculated phase offsets have a non-zero slope as shown in Fig. 3-3(b), the image is resampled after accounting for the irregularity so that the phase offsets would have zero slope as shown in Fig. 3-3(a). This feedback is used in the receive chain in Fig. 3-4 where the output of the 2D-FFT is shown to be used to calculate the sampling offsets which is then fed back to the corner estimation block. The remaining blocks in the receive chain shown in Fig. 3-4 are the counterparts of the ones in the transmit chain shown in Fig. 3-1.

Chapter 4

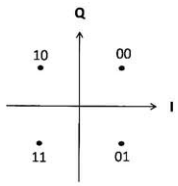
PixNet's Transmitter

Analogous to the way that a standard OFDM transmitter translates bits into voltages, the PixNet transmitter translates bits into pixel intensity values on the LCD. There are two important differences however. First, RF communication uses high frequency carrier waves to transmit both real and imaginary components whereas pixel intensity values are purely real and support no imaginary components. Second, traditional OFDM is one dimensional but in contrast, LCD screens are two dimensional (three dimensional including the time dimension).

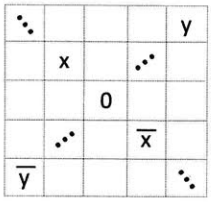
To handle these differences, PixNet generalizes the standard OFDM algorithm by making some key modifications. In the modified algorithm, bits are first error protected, modulated into complex numbers and then broken down into symbols just like standard OFDM. However, instead of feeding these into a one dimensional IFFT, each symbol is first arranged into a two dimensional Hermitian matrix. This matrix is then fed into a 2D IFFT. The special properties of the Hermitian arrangement ensure that the output from the IFFT is entirely real. Additionally, by using a 2D IFFT we also get a two dimensional output appropriate for display on an LCD. To actually transmit the data, the transmitter stacks in each frame as many OFDM symbols as the LCD can support. It then transmits frame after frame until the transfer is complete.

4.1 Modulation

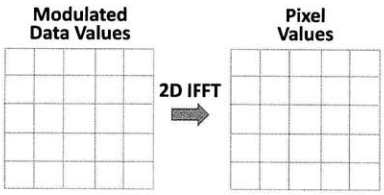
PixNet uses 4QAM modulation which produces a complex number for every pair of bits. Fig. 4-1(a) illustrates this process and shows the four complex numbers that correspond to all possible values of a



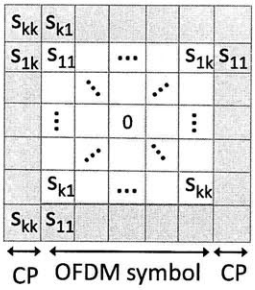
(a) 4QAM maps a bit pair to one of four complex numbers



(b) A Hermitian OFDM symbol



(c) IFFT maps modulated bits to pixel values



(d) An OFDM symbol with a Cyclic Prefix (CP)

Figure 4-1—Components of PixNet's Transmitter.

sequence of 2 bits.

4.2 Maintaining Real Output

As mentioned above, PixNet ensures that the output of the transmitter is purely real by taking advantage of the intrinsic property of Fourier transforms, which says that if the input function is Hermitian, its IFFT is real.[1] To turn the modulated bits into Hermitian matrices, PixNet first breaks the stream of modulated values into symbols. For example, say we have a transmitter block of 5×5 pixels. We can use it to transmit 12 real data values as follows: We first organize the values into a Hermitian form, i.e., we assign to pixel $(-i,-j)$ the complex conjugate of the value assigned to pixel (i,j), as shown in Fig. 4-1(b). We then apply a 2D IFFT. The resulting 25 values are all real and can be as pixels' luminance. We refer to such a 2D block of encoded pixels as a 2D OFDM symbol.

One may think that, by transmitting only 12 complex numbers in 25 real values, we have reduced the

[1] A Hermitian function is a complex function with the property that its complex conjugate is equal to the original function with the variable changed in sign, i.e., $f(-x) = \overline{f(x)}$.

efficiency to 50%. Recall however that each complex number is composed of two real numbers. Hence, 12 complex numbers already have 24 real values. As for the center frequency, OFDM naturally does not transmit data in the DC frequency. Furthermore, in PixNet the DC frequency is the average pixel luminance (by definition of IFFT). Upon reception, the average pixel luminance typically gets badly distorted due to the ambient light in the environment. PixNet filters the distortion caused by ambient light by simply ignoring the DC frequency.

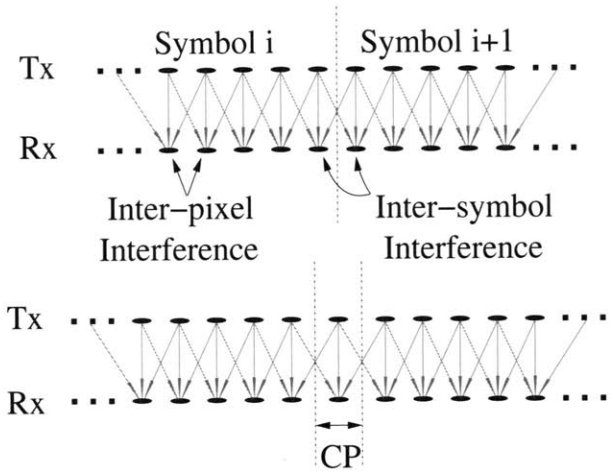


Figure 4-2—Inter-pixel and Inter-symbol interference encountered in the LCD-camera channels. Each pixel at the transmitter is shown to interfere with its neighboring pixel at the receiver. After the introduction of the cyclic prefix, the Inter-symbol interference disappears.

4.3 Lack of pixel-to-pixel matching between the LCD and Camera

In addition to modifying OFDM to meet the special requirements imposed by the LCD-camera channel, PixNet takes advantage of the standard properties of OFDM to solve problems created by the LCD-camera channel. Specifically, there is no one-to-one mapping between the pixels on the LCD and the pixels on the camera. Each pixel on the LCD is a light source and light is diffusive in nature. The camera's lens attempts to concentrate the light from each LCD pixel onto a small region in the image. In practice, focus can never be perfect and hence, each camera pixel receives light from multiple nearby LCD pixels as shown in Fig. 4-2.[2] As a result, the transmitter pixels bleed into each other at the receiver. This bleeding of the pixels creates two effects on the received samples: First, pixels are subject to interference from pixels belonging

[2]The pixels on the camera (Rx) and the LCD (Tx) are shown to be spatially aligned only for demonstration purposes. Typically, the number of pixels on the camera and the LCD are different and the pixels are not aligned.

to neighboring symbols (inter-symbol interference). Second, pixels from the same symbol interfere with each other creating a blur effect (inter-pixel interference).

The above problem is analogous to multipath in RF wireless channels. Specifically, in environments with multipath, multiple copies of the signal traverse different paths and arrive at the receiver with different delays. As a result, each received signal sample is a linear combination of a few consecutive transmitted samples. RF transmitters deal with multipath by employing OFDM and appending a *cyclic prefix* to their signal [7]. We adopt a similar approach in PixNet. Adding a cyclic prefix (CP) means repeating the first few signal samples at the end of the symbol [7]. In PixNet, we append the CP around the entire symbol, as shown in Figure 4-1(d). Specifically, we copy the first few rows from the top of the symbol to the bottom of the symbol and vice versa. We also do the same for a few columns at the left and right edges of the symbol. Since the cyclic prefix introduces a separation between the two symbols as shown in Fig 4-2, it prevents pixels belonging to different symbols from interfering with one another. Inter-pixel interference is dealt with by the fact that OFDM transforms the samples into the frequency domain. The seemingly complex operation of pixel bleeding which is a convolution in the spatial domain turns into a simple multiplication operation in the frequency domain. This makes it easy to address blur from inter-pixel interference as described in §6. The operator can pick the size of the OFDM symbol and the cyclic prefix length. The cyclic prefix length chosen should be greater than the amount by which pixels bleed into each other. Empirically, we found that an OFDM symbol of 81×81 pixels and a cyclic prefix of 5 pixels (introduced on all sides of the symbol) works reasonably well.

4.4 Scaling and Quantization

Before the data can be displayed on an LCD screen, it has to be appropriately scaled and quantized to match the 8-bit resolution that current displays support. The output of a 2D IFFT has a very high Peak to Average Power Ratio (PAPR). And hence, a straightforward 8-bit quantization results in very poor resolution. To increase the resolution at the display, PixNet clips the output of the 2D IFFT before quantizing it. Though clipping increases resolution, it reduces the SNR of the signal. Therefore, a threshold for clipping is chosen such that the SNR at the transmitter is at least 15 dB, which is acceptable for communicating 4QAM modulated signals over visible light channels.

The final frame to be diplayed is formed by stacking as many 2D OFDM symbols as can fit on the

screen. A single frame was comprised of a total of 20×12 2D OFDM symbols. A small colored border is appended around the stacked symbols to enable corner detectection at the receiver.

Chapter 5

PixNet's Receiver

PixNet's receiver works by first extracting the LCD's coded frame from the frame captured by the camera, using a corner detection algorithm. It then corrects for perspective distortion present in this extracted frame and obtains the right samples to be passed through an FFT for normal OFDM decoding.

5.1 Corner Detection

Before the receiver can start decoding, it needs to separate the coded frame from the background in the image. This requires the receiver to detect the four corners of the coded frame. Corner detection is a widely studied problem [21]. In particular, the literature on 2D barcodes has a variety of corner detection algorithms. PixNet can work with any of them. Our implementation uses the Data Matrix corner detection algorithm described in [3].

5.2 Perspective Correction

When a camera takes a picture of an object, it projects that object on to the plane of its sensors. This projection includes shifting the object with respect to its surroundings, scaling the object, and in general, distorting the geometry of the original object. This problem of perspective transformation is widely studied in computer vision [5]. Our context however differs from the traditional context studied in computer vision. On the one hand, our constraints are stricter since we cannot use offline algorithms or tolerate minor distortions which are typically acceptable to the human eye [5]. On the other hand, we have more flexibility

in our design since we can encode the imaged object in a manner that simplifies correcting for perspective distortion. Because of these differences, we do not use traditional perspective correction algorithms. Instead we develop our own algorithm which generalizes the OFDM sampling correction algorithm.

The intuition underlying our approach is simple: We approach perspective distortion as a sampling problem. Specifically, the LCD pixel values refer to the signal samples at the transmitter. The camera pixel values refer to the signal samples at the receiver. When the LCD and camera are at an angle, some parts of the LCD are closer to the camera, and hence occupy a relatively bigger space in the image. This emulates the effect of sampling these parts at a higher rate. Parts of the LCD that are further away from the camera, occupy a relatively smaller space in the image, and hence it is as if they were sampled at a slower rate. To correct for perspective distortion, PixNet needs to find the relationship between the sampling points on the camera and those on the LCD, and re-sample the received image at the locations that best reflect the original LCD samples.

Sampling differences between sender and receiver occur in RF channels too. The DAC at the transmitter and the ADC at the receiver are not synchronized to sample the signal at the same instants. Furthermore, they typically differ in their sampling frequency. As a result, the receiver's samples are different from the transmitter's samples. To decode the signal properly, the receiver needs to resample the signal as closely as possible to the transmitter's samples. However, in RF channels, the sampling is regular and hence there is no analogy to the case of geometric distortion.

In this section, we build on this intuition to provide a generalized OFDM sampling correction algorithm that works with perspective distortion. Before dealing with the general case of irregular sampling, we motivate our approach with a simpler special case, where the samples at the receiver are all shifted by the same amount with respect to the samples at the transmitter.

5.2.1 A Constant Shift Offset

Consider a 2D OFDM symbol sampled with shifts Δ_x on the x-axis, and Δ_y on the y-axis, as shown in Fig. 5.2.1. A basic property of the Fourier transform is that shifts in the signal domain translate into phase offsets in the frequency domain [7].[1] Given this property, it is relatively simple to figure out how a constant sampling shift effects the encoded data. Specifically, since a 2D OFDM symbol is generated by taking a

[1]This argument is subject to the shift being less than the cyclic prefix of the OFDM symbol since any larger shifts cause interference from nearby OFDM symbols.

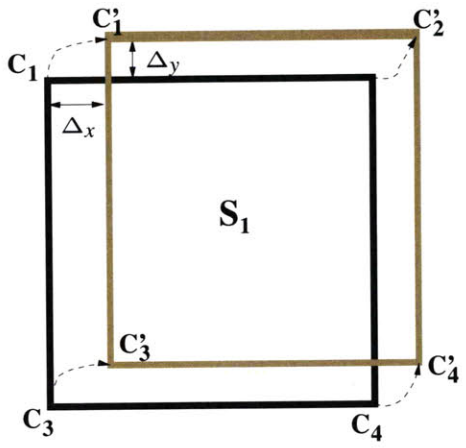


Figure 5-1—A 2D OFDM symbol with shift offsets. The corners were detected incorrectly with an x offset of Δ_x and a y offset of Δ_y

2D IFFT of the complex numbers, $s_{k,l}$, that represent the modulated bits, a sampling shift introduces phase shifts in these complex numbers as follows:

Proposition 5.2.1 *Let $\Delta\theta_x$ be the difference between the phase shifts experienced by $s_{k,l}$ and $s_{k',l}$, and $\Delta\theta_y$ the difference between the phase shifts experienced by $s_{k,l}$ and $s_{k,l'}$. Then:*

$$(\Delta\theta_x, \Delta\theta_y) = \left(\frac{2\pi(k-k')\Delta_x}{L_s}, \frac{2\pi(l-l')\Delta_y}{L_s}\right) \tag{5.1}$$

where $L_s \times L_s$ is the size of the 2D-FFT.

Proof The proof is relatively simple. Specifically, the 2D OFDM symbol can be expressed as:

$$S_{m,n} = \sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1} s_{k,l}\, e^{j\frac{2\pi km}{L_s}}\, e^{j\frac{2\pi ln}{L_s}}$$

If The receiver samples the OFDM symbol with a shift (Δ_x, Δ_y), the resulting samples are given by

$$S'_{m,n} = \sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1} s_{k,l}\, e^{j\frac{2\pi k(m+\Delta_x)}{L_s}}\, e^{j\frac{2\pi l(n+\Delta_y)}{L_s}}$$

To decode the OFDM symbol from these samples, the receiver takes a 2D FFT. Given however that the samples are shifted, the FFT does not reproduce exactly the original complex numbers; it produces a phase

shifted version $z_{p,q}$ of the original complex numbers as follows:

$$\begin{aligned}
z_{p,q} &= \frac{1}{L_s^2} \sum_{m=0}^{L_s-1} \sum_{n=0}^{L_s-1} S'_{m,n}\, e^{-j\frac{2\pi mp}{L_s}} e^{-j\frac{2\pi nq}{L_s}} \\
&= \frac{1}{L_s^2} \sum_{m=0}^{L_s-1} \sum_{n=0}^{L_s-1} \sum_{k=0}^{L_s-1} \sum_{l=0}^{L_s-1} s_{k,l}\, e^{j\frac{2\pi m(k-p)}{L_s}} e^{j\frac{2\pi n(l-q)}{L_s}} e^{j\frac{2\pi k\Delta_x}{L_s}} e^{j\frac{2\pi l\Delta_y}{L_s}} \\
&= \frac{1}{L_s^2} \sum_{k=0}^{L_s-1} \sum_{l=0}^{L_s-1} s_{k,l}\, e^{j\frac{2\pi k\Delta_x}{L_s}} e^{j\frac{2\pi l\Delta_y}{L_s}} \sum_{m=0}^{L_s-1} e^{j\frac{2\pi m(k-p)}{L_s}} \sum_{n=0}^{L_s-1} e^{j\frac{2\pi n(l-q)}{L_s}} \\
&= \sum_{k=0}^{L_s-1} \sum_{l=0}^{L_s-1} s_{k,l}\, e^{j\frac{2\pi k\Delta_x}{L_s}} e^{j\frac{2\pi l\Delta_y}{L_s}} \delta(k-p)\delta(l-q) \\
&= s_{p,q}\, e^{j\frac{2\pi p\Delta_x}{L_s}} e^{j\frac{2\pi q\Delta_y}{L_s}}
\end{aligned}$$

Thus, each complex number, $s_{k,l}$ experiences a phase shift $\frac{2\pi k\Delta_x}{L_s} + \frac{2\pi l\Delta_y}{L_s}$. Thus, the difference in phase shift between $s_{k,l}$ and $s_{k',l}$ is $\frac{2\pi(k-k')\Delta_x}{L_s}$ and that between $s_{k,l}$ and $s_{k,l'}$ is $\frac{2\pi(l-l')\Delta_y}{L_s}$.

The above proposition allows the receiver to estimate the sampling shifts (Δ_x, Δ_y) from differences in the phase shifts of the modulated complex numbers, $(\Delta\theta_x, \Delta\theta_y)$. But how does the receiver obtain $(\Delta\theta_x, \Delta\theta_y)$? To enable the receiver to compute $(\Delta\theta_x, \Delta\theta_y)$, we transmit a few known complex numbers in each OFDM symbol, which we refer to as *pilots*. Since the receiver knows the original phase of each transmitted pilot, it can easily compute the phase shift experienced by each pilot. The receiver then computes differences in phase shifts between two pilots in the same row, $\Delta\theta_x$, and two pilots in the same column $\Delta\theta_y$, and substitutes them in eq. 5.1 to obtain the sampling shifts. Once it has the sampling shifts, it resamples the OFDM symbol at the correct locations, and computes the FFT again to obtain the correct modulated complex numbers. Finally the receiver demodulates the complex numbers to obtain the bits using standard 4QAM demodulation.

5.2.2 General Offsets

We now consider a general case where we sample a 2D OFDM symbol with a perspective distortion as shown in Fig. 5-2. Following from the observation above, we suspect that these corner shifts should similarly result in phase offsets at the receiver.

However, unlike the previous case where we needed to estimate 2 unknowns (Δ_x, Δ_y), in this case

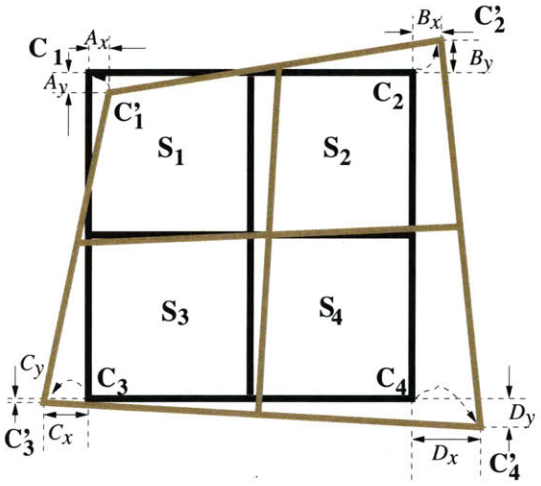


Figure 5-2—A set of four 2D OFDM symbols with perspective distortion. The corners C_1, C_2, C_3 and C_4 are detected with x and y corner offsets of (A_x, A_y), (B_x, B_y), (C_x, C_y) and (D_x, D_y) respectively.

we need to estimate 8 unknowns (A_x, A_y), (B_x, B_y), (C_x, C_y) and (D_x, D_y). If we consider any particular symbol we would have only 2 equations, one for phase shifts θ_x, and another for θ_y. But we need at least 8 equations to estimate these 8 offsets. We solve this problem by considering a super-symbol consisting of 4 symbols as shown in Fig. 5-2. Now, we can generate 2 equations for each symbol, resulting in 8 equations overall. In the appendix, we prove the following:

Proposition 5.2.2 *Consider a* 2×2 *2D OFDM super-symbol generated by taking the IFFT of the complex numbers* $s_{k,l,r}$*, where* k *and* l *are integers between* 0 *and* $L_s - 1$ *and* $r \in \{1,2,3,4\}$ *denotes the symbol index. Let us assume that this symbol is sampled with relatively small x and y corner offsets of* (A_x, A_y), (B_x, B_y), (C_x, C_y) *and* (D_x, D_y) *at its four corners, as show in Fig. 5-2. Let* $\Delta\theta_{x,r}$ *be the difference between the phase shifts experienced by* $s_{k,l,r}$ *and* $s_{k',l,r}$*, and* $\Delta\theta_{y,r}$ *be the difference between the phase shifts experienced by* $s_{k,l,r}$ *and* $s_{k,l',r}$*. Then:*

$$\begin{pmatrix} \Delta\theta_{1,x} \\ \Delta\theta_{2,x} \\ \Delta\theta_{3,x} \\ \Delta\theta_{4,x} \end{pmatrix} = \frac{2\pi(k-k')}{16L_s} \begin{pmatrix} 9 & 3 & 3 & 1 \\ 3 & 9 & 1 & 3 \\ 3 & 1 & 9 & 3 \\ 1 & 3 & 3 & 9 \end{pmatrix} \begin{pmatrix} A_x \\ B_x \\ C_x \\ D_x \end{pmatrix}$$

$$\begin{pmatrix} \Delta\theta_{1,y} \\ \Delta\theta_{2,y} \\ \Delta\theta_{3,y} \\ \Delta\theta_{4,y} \end{pmatrix} = \frac{2\pi(l-l')}{16L_s} \begin{pmatrix} 9 & 3 & 3 & 1 \\ 3 & 9 & 1 & 3 \\ 3 & 1 & 9 & 3 \\ 1 & 3 & 3 & 9 \end{pmatrix} \begin{pmatrix} A_y \\ B_y \\ C_y \\ D_y \end{pmatrix} \tag{5.2}$$

PROOF. Refer to the Appendix.

Note that eq. 5.2 is a generalized version of eq. 5.1 for $(A_x, A_y) = (B_x, B_y) = (C_x, C_y) = (D_x, D_y) = (\Delta_x, \Delta_y)$. After estimating the offsets (A_x, B_x, C_x, D_x), (A_y, B_y, C_y, D_y) with the help of eq. 5.2 using the pilot bins as described in sec. 5.2.1, the receiver re-samples each OFDM symbol at the correct sampling points and computes the FFT again. The resulting complex numbers are demodulated to obtain the transmitted bits.

5.3 Offset Correction Algorithm

At the receiver, PixNet divides all the symbols present in a frame into blocks of 2×2 symbols called super-symbols as shown in Fig. 5-2. Using the detected four-corners of the frame, corners are estimated for each of the super-symbols by assuming an uniform arrangement of 20×12 symbols per frame. Since the original four corners themselves are inaccurate, the estimated corners for each of the super-symbols will be inaccurate as shown in Fig. 5-2. In what follows, we describe how the corners are corrected using the properties of the Fourier Transform.

The receiver samples each super-symbol by applying a uniform grid using the estimated corners for the respective super-symbols. The samples for each of the four symbols of a super-symbol are then passed as input to a 2D FFT block. The output of the 2D FFT would have been the actual modulated data which was transmitted if there were no corner offsets and everything else remained ideal. However, the presence of

corner offsets results in phase offsets between the transmitted and received samples. If, $\theta_{r,x}, r \in \{1,2,3,4\}$ are the x phase offsets and similarly $\theta_{r,y}, r \in \{1,2,3,4\}$ are the y phase offsets, then the relationship between them and corner offsets A_x, B_x, C_x, D_x and A_y, B_y, C_y, D_y is given by equations (5.2).

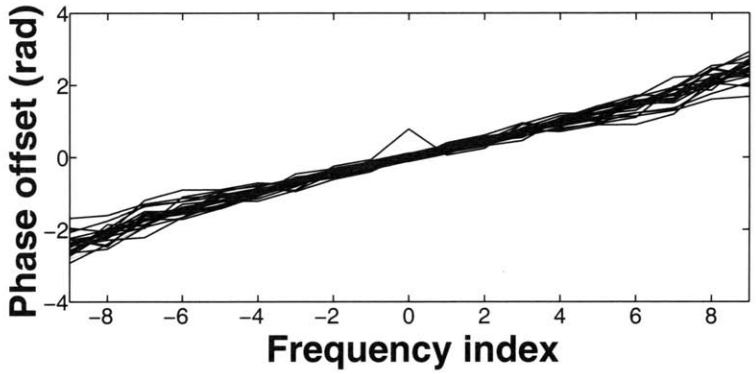


Figure 5-3—Phase offsets for a symbol with non zero corner offsets. We find that all the y frequencies have the same x phase offsets and similarly, all the x frequencies have the same y phase offsets.

Shown in Fig 5-3 are the observed x phase offsets as a function of x frequency for different y frequencies of a symbol. By transmitting known data in some of the frequencies (called as pilot frequencies), these phase offsets are estimated for each of the four symbols in a super-symbol. Specifically, PixNet uses the x and y frequencies $\{-14,-9,-4,-1,1,4,9\}$ as pilots and computes the respective phase offsets for each symbol in a super symbol by taking the mean of 1d-regressions over these frequencies. After computing the phase offsets, corner offsets are estimated using the equations (5.2). The corners for each of the super-symbol are then corrected based on the estimated corner offsets and a new sampling grid based on the corrected corners is now used to sample the super-symbol. The resulting phase offsets after correcting for the corner offsets are now almost zero as shown in Fig. 5-4, implying that current sampling is optimal.

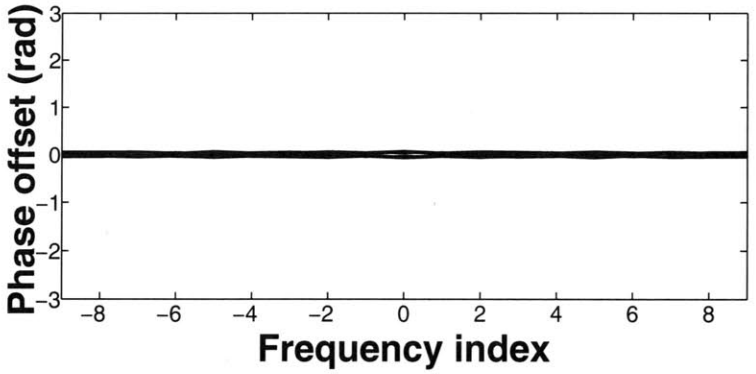


Figure 5-4—Phase offsets for a symbol with zero corner offsets. Both the x and y frequencies exhibit zero phase offsets when the sampling is optimal.

In order to remain robust to variations caused by noise present in the image, this process of sampling, estimating and correcting the corner offsets is repeated twice for each super-symbol. The output of the 2D-FFT for each symbol is then considered as received modulated data which is further demodulated and decoded for error protection. The various blocks that make up PixNet's receive chain are shown below in Fig. 5-5

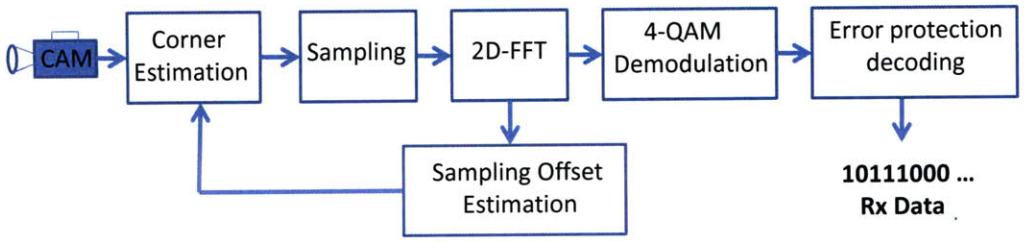


Figure 5-5—PixNet's Receive chain.

Chapter 6

Blur Adaptive Coding

In this chapter, we describe how PixNet deals with blur which is a result of inter-pixel interference, lack of focus or movement while capturing the images.

Blur eliminates sharp transitions in an image and causes nearby pixels to blend together. Thus, its impact is similar to that of a low pass filter, i.e., it attenuates the high frequencies in an image. Fig. 6-1 shows the signal amplitude as a function of the x frequency index (f_x) for a transmitted signal and its received version. The transmitted signal is chosen to have the same energy in all frequencies. We can see from the figure that the high frequencies (i.e., those above 20 or below -20) are heavily attenuated. Thus, these frequencies cannot be used for transmitting information. Conversely, the lowest frequencies (between -10 and 10) are mostly preserved and can deliver information with almost no error. Finally, the frequencies in between, experience significant attenuation but can still be used to transmit some information.

Since PixNet operates in the frequency domain, it can naturally deal with different frequencies experiencing different attenuation. PixNet completely suppresses very high frequencies, and does not use them to transmit information. Frequencies that experience low or moderate attenuation are used to transmit information, but protected with a Reed Solomon error correcting code. An error correction code is chosen for each frequency with redundancy commensurate with the attenuation that frequency experiences.

We find that the amount of attenuation varies not only with the frequency, but it also varies with the position of the symbol in a given frame. Specifically, cameras focus on a particular plane. Objects that are either nearer or farther from the focus plane experience blur (this is typically known as limited depth-of-field (DoF)) [24]. Thus, when the LCD and camera have a view angle with respect to each other, only the center of the LCD will be in focus. Symbols away from the center of the frame are not in the plane

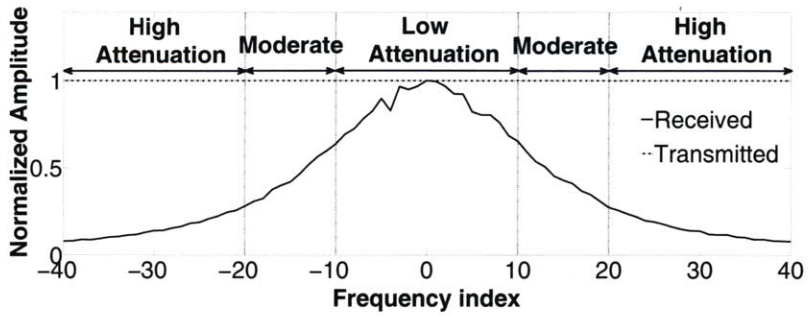


Figure 6-1—Received signal amplitude across different frequencies with a DELL 30" flat display and a Nikon D3X camera at a distance of 2 meters. This plot reveals that the amount of attenuation in the received signal increases with frequency and not all spatial frequencies are actually supported by the LCD-Camera channel. Especially, the high frequencies are very heavily attenuated are not useful for transmission.

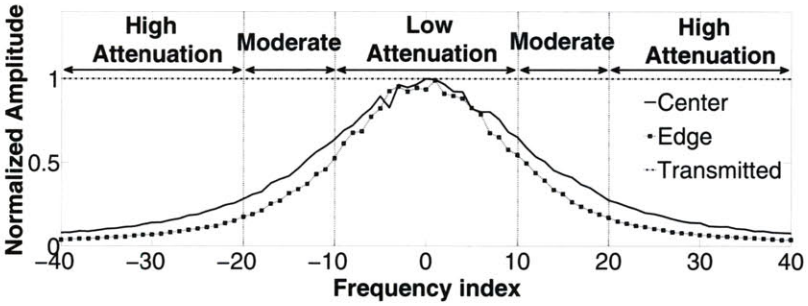


Figure 6-2—Received signal amplitude across different frequencies for symbols that are in the center of the frame and those that are in the edges of the frame. We find that there is a considerable difference in the amount of attenuation experienced by symbols that are at the center of the frame compared to those that are at the edges. Hence, different symbols are encoded for error protection with different amounts of redundancy.

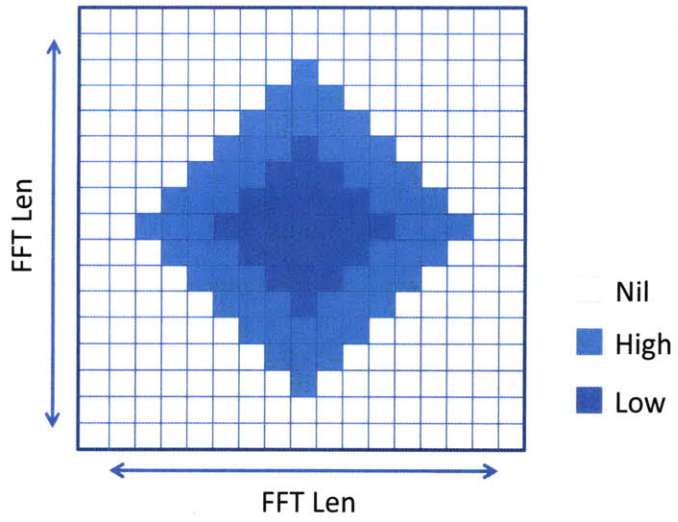


Figure 6-3—The amount of redundancy introduced for RS codes within a symbol as a function of the frequency index (f_x, f_y)**.** Different blocks within a symbol are encoded with different amounts of redundancy depending on their frequencies.

of focus and hence experience increased attenuation due to blur (in addition to perspective distortion) as shown in Fig. 6-2. PixNet exploits this information while optimizing the redundancy in the error correcting code. Specifically, the symbols at the center of the frame will be encoded with lower redundancy and the symbols away from the center of the frame will be encoded with higher redundancy.

The configuration of the Reed Solomon code in PixNet is as follows: The code operates on a block size of 255, where each element in the block is 8 bits [6]. Based on empirical calculations, we decide to use codes that correct up to 3% (weak), 7% (medium), and 15% (strong) errors. Frequencies such that $|f_x| + |f_y| \leq 10$ are always coded with the weak code as shown in Fig. 6-3. Frequencies such that $20 \leq |f_x| + |f_y| < 10$ are coded with the medium code for symbols near the center of the frame and with the strong code for symbols away from the center of the frame. Frequencies such that $40 \leq |f_x| + |f_y| < 20$ are not used for transmitting any information and hence are not error protected.

Chapter 7

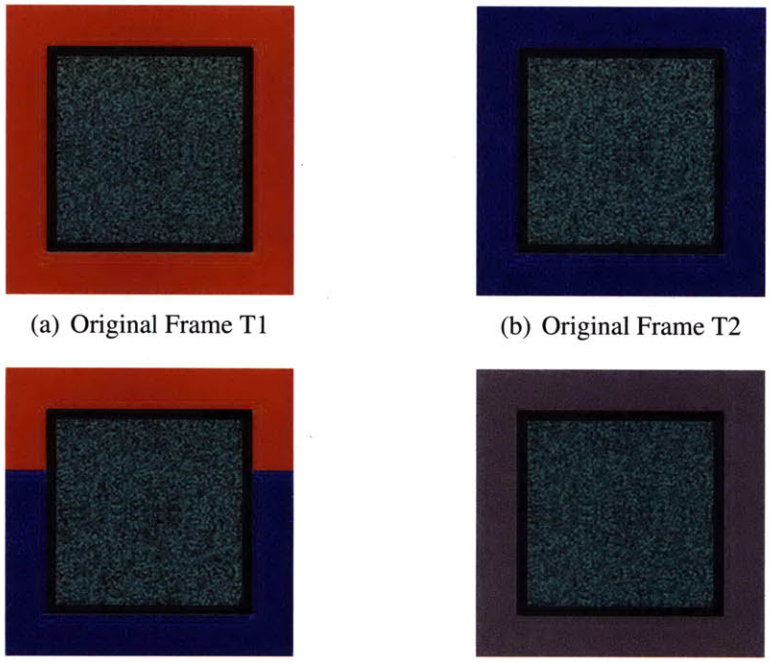
Frame Synchronization

Unsynchronized operation between LCD and camera can cause the camera to capture an image while the LCD is in the process of rendering a new frame. We have performed detailed experiments to identify specific distortions arising from the lack of transmitter-receiver synchronization. We observed the following two distortions:

- *Frame Splitting*: This distortion causes two consecutive LCD frames to form the top and bottom parts of one frame captured by the camera as shown in Fig 7-1(c). This occurs because the graphics card at the transmitter writes the next frame to the LCD's frame buffer while the LCD has not completely rendered the current frame. This is the so-called 'vertical sync' problem that occurs when the copying of frames by the graphics card is out of sync with the rendering of frames by the LCD screen.
- *Frame Shadowing*: In this case, the image captured by the camera is a linear combination of two consecutive LCD frames as shown in Fig 7-1(d). Specifically, the camera captures light for the duration that its shutter remains open, and integrates over this time period. Thus, if the LCD moves from rendering one frame to the next while the camera's shutter is open, the camera integrates light from both LCD frames. The net effect of this integration process is a shadowed frame.

7.1 Frame Splitting

PixNet tackles these problems as follows. For frame splitting, we leverage the OpenGL graphics library [20]. OpenGL provides the `GLFlush` and `GLFinish` interface commands as part of its API to



(a) Original Frame T1

(b) Original Frame T2

(c) A frame split between T1 and T2

(d) A shadowed frame containing both T1 and T2

Figure 7-1—Frame splitting and frame shadowing observed at the receiver. Frame splitting occurs because of the lack of synchronization between the graphics card and the display. Frame shadowing on the other hand, occurs because of the lack of synchronization between the LCD and the camera.

force the graphics card to wait for the LCD screen to complete rendering of the previously copied frame before the next frame is copied to its frame buffer. This effectively eliminates the frame splitting problem.

7.2 Frame Shadowing

We calibrate the shutter speed to ensure that the camera's frame rate is at least twice that of the LCD. This configuration ensures that each frame displayed by the LCD is captured by the camera without shadowing at least once. Refer to Fig. 7-2.

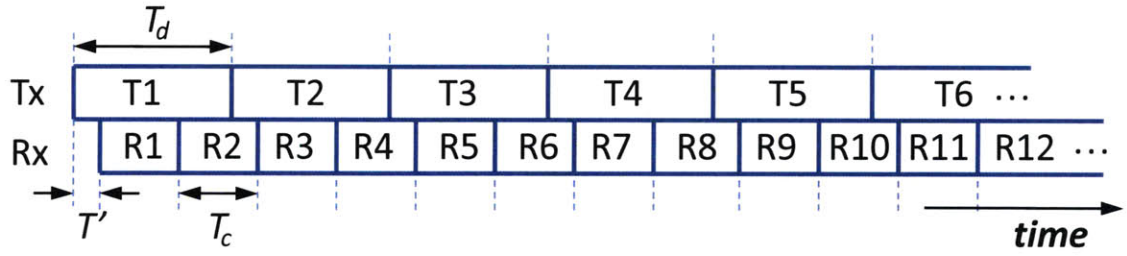


Figure 7-2—Transmitter and receiver timing diagram. The camera (Rx) is operating at twice the speed of the LCD (Tx). $T_d = 2T_c$. We find that when $T_d \geq 2T_c$, every transmitted frame is received at the receiver without shadowing at least once.

Let us assume that at the transmitter, each frame Ti (T1, T2, T3, ...) is displayed during the time interval $((i-1)T_d, iT_d)$ (Typically T_d is either 16.67 ms or 33.33 ms for a 60Hz and a 30Hz display respectively). And let the camera capture frames Ri (R1, R2, R3, ...) in the intervals $(T' + (i-1)T_c, T' + iT_c)$, where T' is some arbitrary misalignment interval due to the lack of synchronization between the LCD and the camera as shown in Fig. 7-2.

The camera integrates light incident on its sensors from the moment the shutter is opened to the moment it is closed (also known as the exposure time T_c). And hence, each image that is captured by the camera is potentially a combination of multiple frames that were displayed during the exposure interval. For ex., in Fig. 7-2, R2 is a combination of the frames T1 and T2 and therefore exhibits shadowing.

However, if $T_d \geq 2T_c$, we observe that every transmitted frame is received without shadowing (inter-frame interference) in at least one of the received frames. For ex., T1 is received without interference in R1, T2 is received without interference in R3, and T3 in R5, etc. PixNet uses this principle in its design by having the LCD refresh frames at 30 fps and having the camera capture frames at 60 fps. Frames containing interference from other frames as in Fig. 7-1(d) are disambiguated from frames not containing interference as in Fig. 7-1(a) and Fig. 7-1(b) by using Red and Blue colors alternately as borders for succeeding frames.

Since the camera measures light intensity in each of the three channels (Red, Green and Blue) separately, non-shadowed frames are identified by looking at the intensity difference between the Red channel and the Blue channel in the border regions of the frame. A very high intensity difference between the Red and Blue channels identifies a non-shadowed frame. A moderately low intensity difference between the Red and Blue channels identifies a shadowed frame and such frames are discarded.

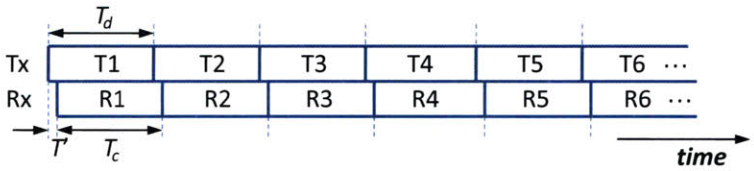
Chapter 8

3D OFDM

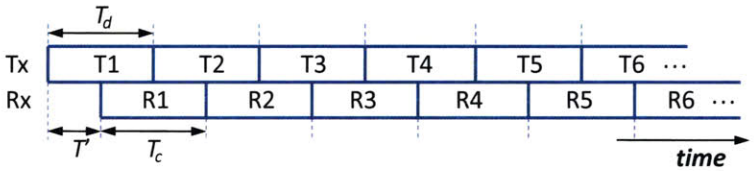
By using 2D FFTs and by designing two-dimensional offset correction algorithms, PixNet generalizes the traditional 1D OFDM into a 2D OFDM system. We now extend this generalization further by designing a 3D OFDM system for the LCD-Camera channel.

PixNet addresses inter-pixel interference (blur) and perspective distortion by operating in the frequency domain. Since both blur and perspective distortion are 2D effects, PixNet used 2D OFDM to deal with them. As described in §7, there is however another interference effect that comes into play while transmitting and receiving frames over the LCD-Camera channel: Inter-frame interference (shadowing) as shown in Fig. 7-1(d). This problem was tackled in §7 by using a lower transmission rate (30 fps) at the LCD. Though using a lower frame rate of 30 fps avoids the problem of shadowing, it reduces the achievable throughput (Since, effectively only 30 frames are communicated every second).

Another way of tackling the problem of inter-frame interference is by coding across the time dimension. This is the motivation for designing a 3D OFDM system. Just as how 2D OFDM helps us deal with inter-pixel interference, we expect 3D OFDM to help us deal with inter-frame interference (along with inter-pixel interference). Since both the transmitter and the receiver can now operate at 60 fps as shown in Fig. 8-1, 3D OFDM can potentially deliver higher throughput. Also note that since the transmitter and the receiver are not synchronized, the amount of inter-frame interference in the received frames can vary across different transmissions as shown in Fig. 8-1. This lack of alignment (denoted by T') depends upon the instant at which the camera's shutter is triggered relative to the instant at which the graphics card at the transmitter starts rendering the frames.



(a) Tx and Rx are almost perfectly synchronized ($T' \sim 0$)

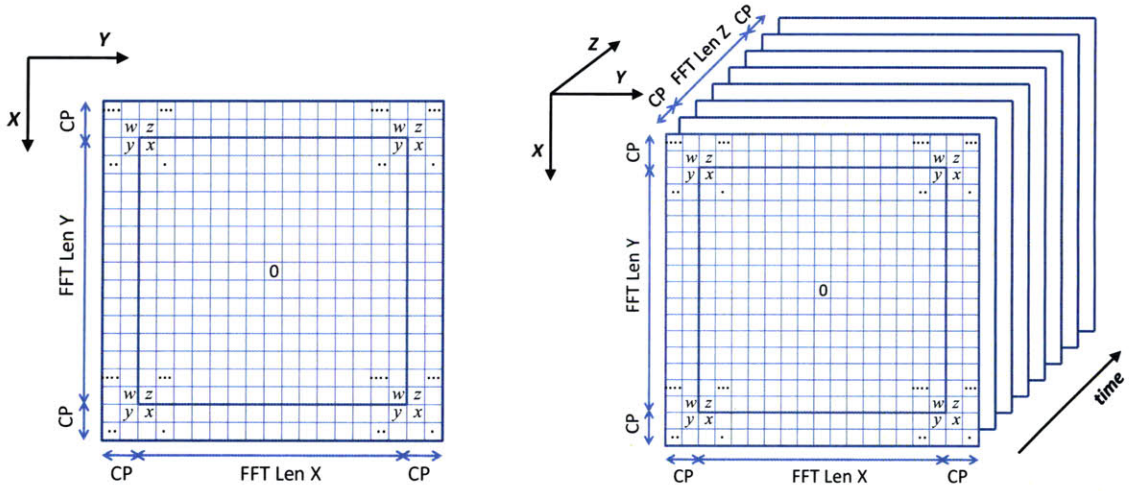


(b) Tx and Rx are completely out of synchronization ($T' = \frac{T_d}{2}$)

Figure 8-1—Transmitter and receiver timing diagram. The Camera (Rx) and LCD (Tx) are operating at the same speed ($T_d = T_c$). The amount of inter-frame interference (shadowing) varies depending on when the camera's shutter is triggered relative to the graphics card at the transmitter.

8.1 Model

We design a 3D OFDM system for PixNet by using the time dimension along with the two spatial dimensions. Refer to Fig. 8-2. Each PixNet symbol is now three-dimensional and spans multiple frames. The 2D symbols used by PixNet are of size 81×81 and the 3D symbols are of size $81 \times 81 \times 5$. In comparison with 2D OFDM, 3D OFDM encodes the same amount of data in any given number of frames.



(a) A 2D OFDM symbol

(b) A 3D OFDM symbol

Figure 8-2—Extending PixNet to three dimensions.

8.2 Transmitter

Though extending a 2D OFDM system to a 3D OFDM system might seem pretty straightforward, there are some issues worth discussing. First, the frequencies are now 3 dimensional and hence, blur adaptive coding is done as a function of $|f_x| + |f_y| + |f_z|$ as opposed to $|f_x| + |f_y|$ as described in §6. After performing 4QAM modulation, the data is now arranged in a Hermitian format according to the equation $f_{x,y,z} = \bar{f}_{-x,-y,-z}$. The data is then passed to a 3D IFFT and the output is now suitably clipped, scaled and quantized to match the 8−bit resolution of the LCD screen. The amount of CP introduced for an OFDM symbol is chosen to be equal to the expected extent of interference from neighboring signal points. Since along the z axis, we expect to see interference from only one succeeding frame, we use a CP of length 1 for the z axis. We continue to use a CP of 5 for the x and y axes. As many symbols as can fit on the screen are stacked together into a series of 7 (FFT Len Z + 2CP) frames and a small colored border is appended around the symbols to enable corner detection.

8.3 Receiver

As perspective distortion is a 2D effect, we leverage the afore developed perspective correction algorithm in §5.3 for the 3D scenario as well. Specifically, the LCD displays a 2D encoded frame with a white border before displaying a burst of 3D encoded frames that have red and blue borders alternately. The receiver detects the 2D encoded frame by looking at the border and uses it to determine the optimal sampling grid as described in §5.3. It then uses this sampling grid to sample all the following 3D encoded frames. The samples are then fed into a 3D FFT. The output of the FFT is demodulated and decoded for error protection.

Chapter 9

Evaluation

In this chapter, we implement a prototype of PixNet and evaluate its performance using commodity LCDs and cameras. We also compare PixNet to QR code, a state-of-the-art 2D barcode.

Hardware: For our transmitter, we use a DELL 30" flat display at a resolution of 2560×1600 pixels. At the receiver end, we experiment with a Casio EX-F1, which has a resolution of 6 megapixels, and a Nikon D3X, which has a resolution of 24 megapixels. In some experiments, we use a Nokia N82 phone, which has a 5 megapixel camera.

Compared Schemes: We compare two encoding schemes:

- **PixNet:** We have implemented a prototype of PixNet in C. All our experiments use a symbol size of 81×81 pixels with a cyclic prefix of 5 pixels.
- **QR codes:** We experiment with a baseline scheme that encodes data using QR code. Our experiments use the open source reference implementation of QR code from [28].

Experimental Setup: Fig. 9-1 illustrates our experimental setup. We fix the location of the LCD and move the camera. We experiment with different distances between the LCD and camera as well as different view angles. We pack as many QR codes or PixNet OFDM symbols that can fit on the screen. We set the camera exposure time to $1/60$ seconds, hence allowing for a frame rate of 60 fps. Since at most half of the frames show shadowing (see §7), the effective frame rate is 30 fps. We use the autofocus functionality of the cameras, except for the blur experiments in which we control the focus to show the impact of different levels of blur.

Metric: We compare PixNet and QR code schemes in terms of their throughput, that is the number of

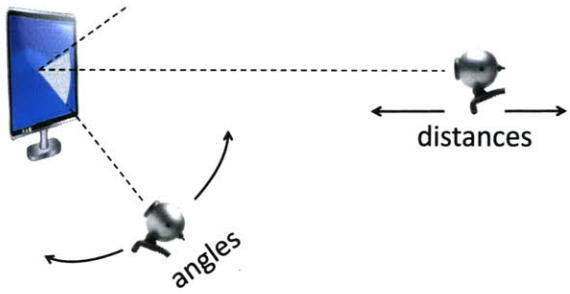


Figure 9-1—Experimental Setup. The camera's location spans different distances and wide view angles while maintaing a line-of-sight to the LCD.

correctly received bits after Reed-Solomon decoding.[1] The throughput is computed as the average per frame throughput multiplied by 30 frames/s. In all experiments we use a shutter speed (exposure time) of 1/60 second, which corresponds to a frame rate of 60 fps and an effective frame rate of 30 fps (Ref. §7.2). The Casio EX-F1 has a burst mode that captures frames at 60 fps. The Nikon D3X FX does not have a burst mode and hence cannot deliver 60 fps in realtime. However by setting the exposure time to 1/60 seconds, we obtain the same quality as if each frame lasted 1/60 second. Our intent however in presenting results for Nikon D3X FX is to show the effect of different sensor types on the performance of PixNet.

Ensuring Fair Comparison: For a fair comparison between PixNet and QR code, we need to ensure that they both have the same upper bound on the amount of information they can pack in a frame. To do so, we note that cameras and LCDs have three color channels: Red, Green, and Blue. QR codes are monochrome and hence set all three channels to the same value. To ensure that PixNet cannot pack more information by using colors, we have PixNet use a single channel, which is the green channel. Also, QR codes are black and white and do not use grey levels. Thus, they allow a maximum of one bit per pixel. PixNet also allows a maximum of one bit per pixel. Specifically, the combination of 4QAM modulation and a hermitian structure results in, on average, one bit per spatial frequency. The number of spatial frequencies is the same as the number of pixels (they are the input and output of IFFT). Thus, both PixNet and QR code can transmit at the same maximum data rates. The actual throughput they achieve in practice depends on how many bits/s they can deliver reliably.

[1]For QR code, we use the default "L" error correction level, which allows for maximum throughput per code.

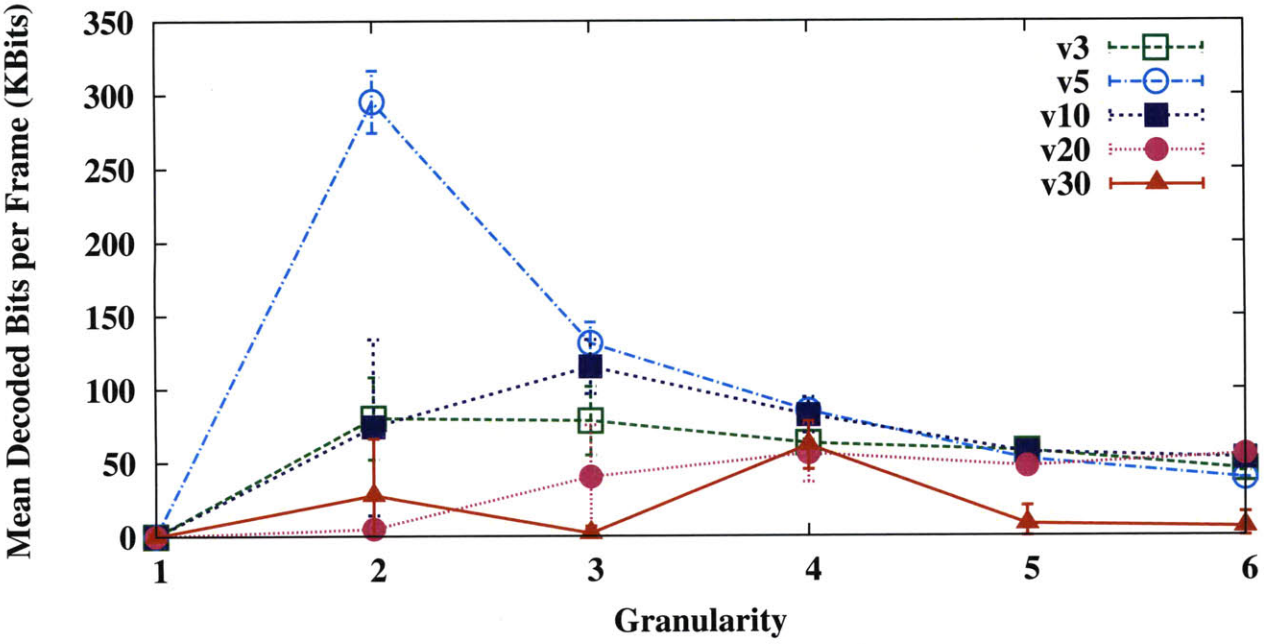


Figure 9-2—Impact of QR code's granularity and version on its performance. QR code **v5** at a granularity of 2 achieves the maximum performance.

9.1 Calibration of QR Code

As mentioned earlier, we allow QR code to transmit one bit/pixel, i.e., to use a granularity of 1. In practice, however, such a granularity suffers excessive errors and reduces the throughput. Thus, we take an additional step and calibrate the QR code parameters for optimal performance. The QR code [2] has two parameters: granularity and version. The granularity defines the number of pixels used to convey a single bit. For instance, a granularity of 2 means that each bit is expressed using 2×2 pixels. The version defines the number of bits encoded in a single QR code. For instance, v5 encodes 864 bits per QR code.

Method. For each version and granularity, we fit as many QR codes as we can on the 30" Dell 3007WFP LCD screen we use in all our experiments. We perform two types of experiments. First, we study the interaction between the version and granularity. To do so, we align the LCD-camera pair (i.e., ensure zero view angle) and fix the distance between them to two meters. We take multiple pictures of the screen using the Nikon D3X camera and decode each frame using the QR code reference implementation [2]. Second, we repeat the above experiment but with different view angles and distances.

Results. Fig. 9-2 shows the effect of different QR code granularities and versions (For clarity, we plot the

subset of the versions that showed better performance). In principle, increasing the version or decreasing the granularity increases the data rate but decreases robustness to errors. The figure shows that QR code version 5 with a granularity of 2 represents a sweet spot in this tradeoff, and yields the best performance. Higher versions pack too many bits in a single QR code, and hence fail to support fine granularities, resulting in lower overall throughput. Lower versions pack too little information in a QR code and hence are not effective at amortizing the per QR code metadata (e.g., corner and synchronization markers).

Fig. 9-3 shows the performance of different QR code versions and granularities as a function of the viewing angle (again, we plot the subset of configurations that showed better performance). It confirms that version 5 at a granularity of 2 performs the best independent of the viewing angle.

We also study the impact of distance and find that there is no clear winner. Specifically, the optimal QR code granularity increases with distance. For the distances that we experiment with (i.e., [2m-16m]) v5 with granularity 2 works well for the shorter distances and version 5 with granularity of 5 works well for the longer distances. Hence, for our distance experiments, we show results for both in §9.2. Finally, note that though we allow QR code to use both configurations to deal with distances, all of the experiments in this work use a single configuration for PixNet (a symbol size of 81×81 pixels with a 5 pixels cyclic prefix).

9.2 Impact of Distance

First, we examine whether an LCD-camera link can deliver data over reasonable indoor distances.

Method. In this experiment, we keep the transmitter and receiver aligned but move the receiver away in increments of a meter, each time allowing the camera to refocus before capturing the frames. This experiment is performed with Nikon and Casio cameras as receivers. For each location, the camera zooms to fill the camera's view with the transmitted frame.

Results. Fig. 9-4 shows the throughput of PixNet and QR code as a function of distance between the transmitter and the receiver. We show results for QR code v5 at granularities of 2 and 5. We observe that in comparison with QR code, PixNet delivers up to $2x - 9x$ higher throughput depending on the distance. We also observe that PixNet continues to achieve throughputs of up to 12 Mbps at a distance of 10m which is representative of a large open space, such as a conference room or auditorium. In contrast, QR code (v5,g2) fails to work beyond 5m. QR code (v5,g5) however is able to support distances comparable to

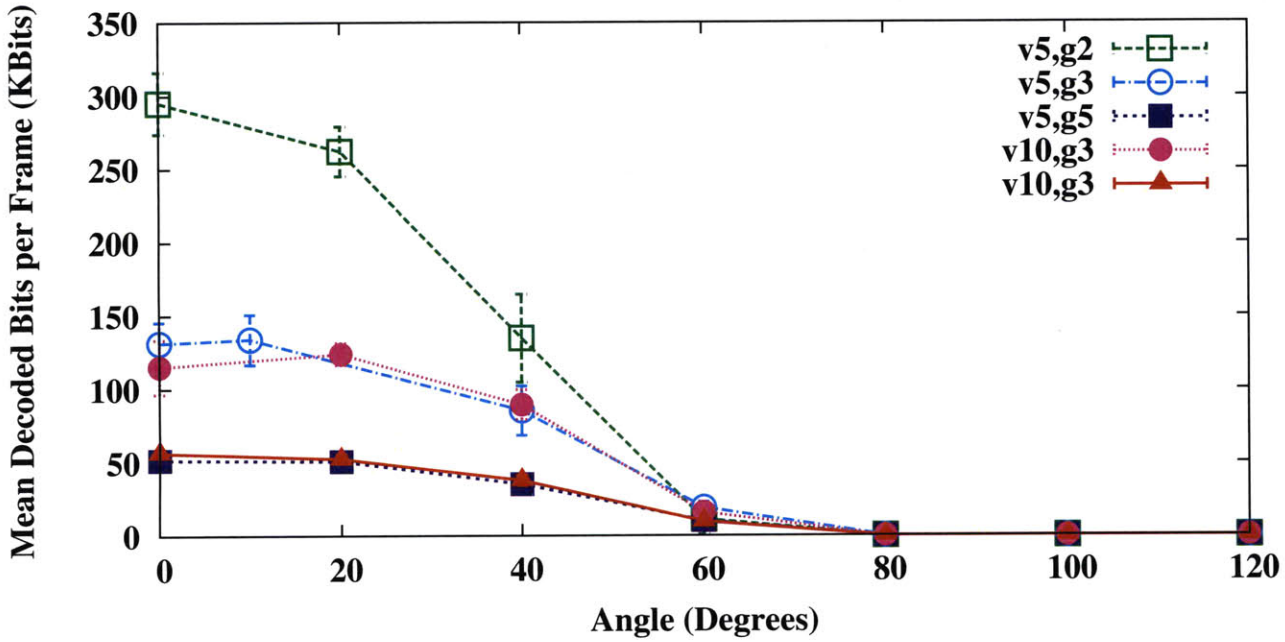


Figure 9-3—Impact of angle on throughput achieved with different QR Code versions and granularities. We find that QR code **v5** at a granularity of 2 has a superior performance for all angles experimented with.

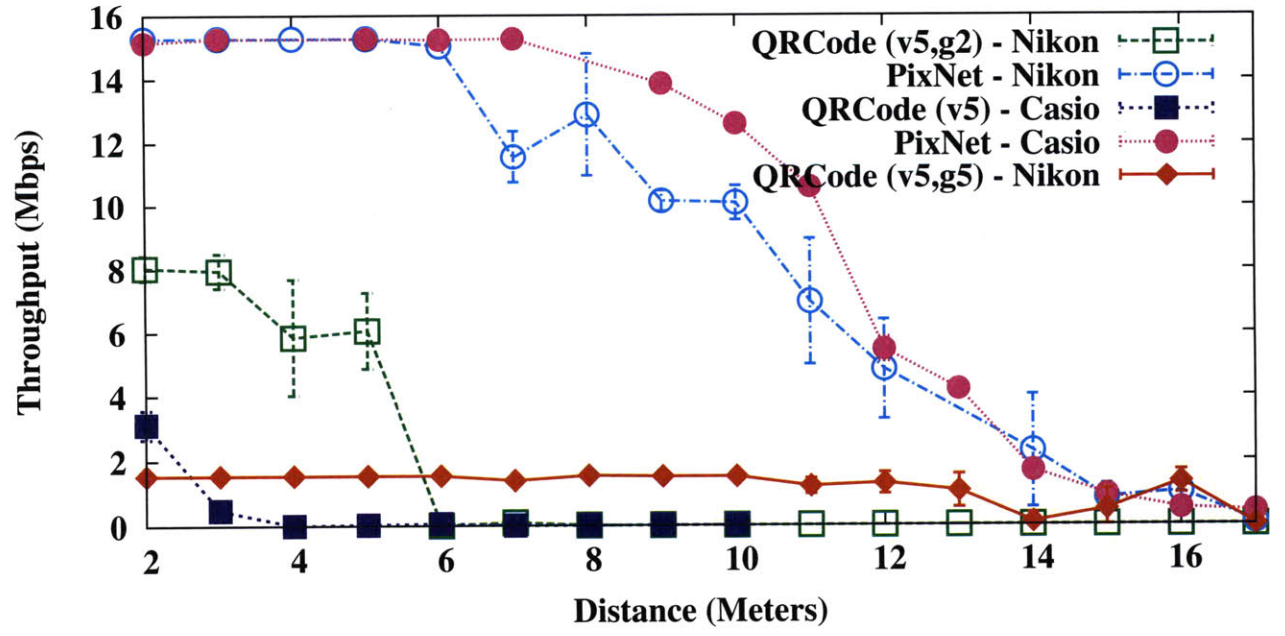


Figure 9-4—Impact of distance on throughput of PixNet and QR Code. We find that PixNet achieves throughputs of up to 12 Mbps at a distance of 10m whereas QR code fails to provide any appreciable throughput beyond 5m. Overall, in comparision with QR code, PixNet's performance is around $2x - 9x$ higher depending on the distance.

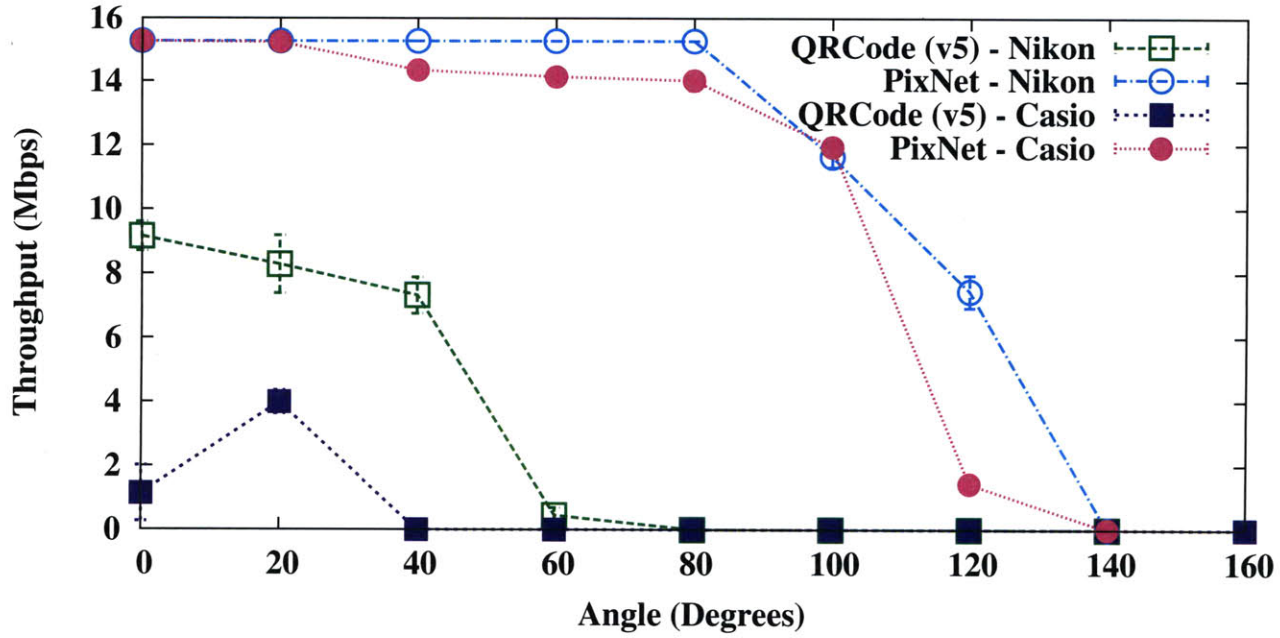


Figure 9-5—Impact of viewing angle on PixNet and QR Code. PixNet has a significantly higher coverage in terms of angles. PixNet continues to deliver 8 Mbps at angles as wide as 120°.

PixNet, but with an order of magnitude worse performance at short distances.

Interestingly, the figure also shows that while PixNet achieves a higher throughput with Casio EX-F1, QR code achieves a higher throughput with Nikon D3X. This is due to the two cameras having different pixel sizes. Nikon D3X has a CCD pixel size of 34.46 μm^2 where as Casio EX-F1 has a CCD pixel size of 193.54 μm^2, which is around 5.6 times bigger than that of Nikon. Having a bigger pixel size allows the pixels to gather more light quickly, and reduces the amount of shot-noise (noise that arises because of small number of photons being incident on the sensor). However, bigger pixel size also means that the light being gathered at each receiver pixel is potentially a combination of light from multiple transmitter pixels, and hence images from Casio have higher blur (inter-pixel interference).

As described in section 6, PixNet has an effective way of dealing with blur. This, combined with the fact that Casio can collect more light makes PixNet perform better with Casio. QR codes on the other hand are highly sensitive to blur and hence perform poorly with Casio when compared with Nikon.

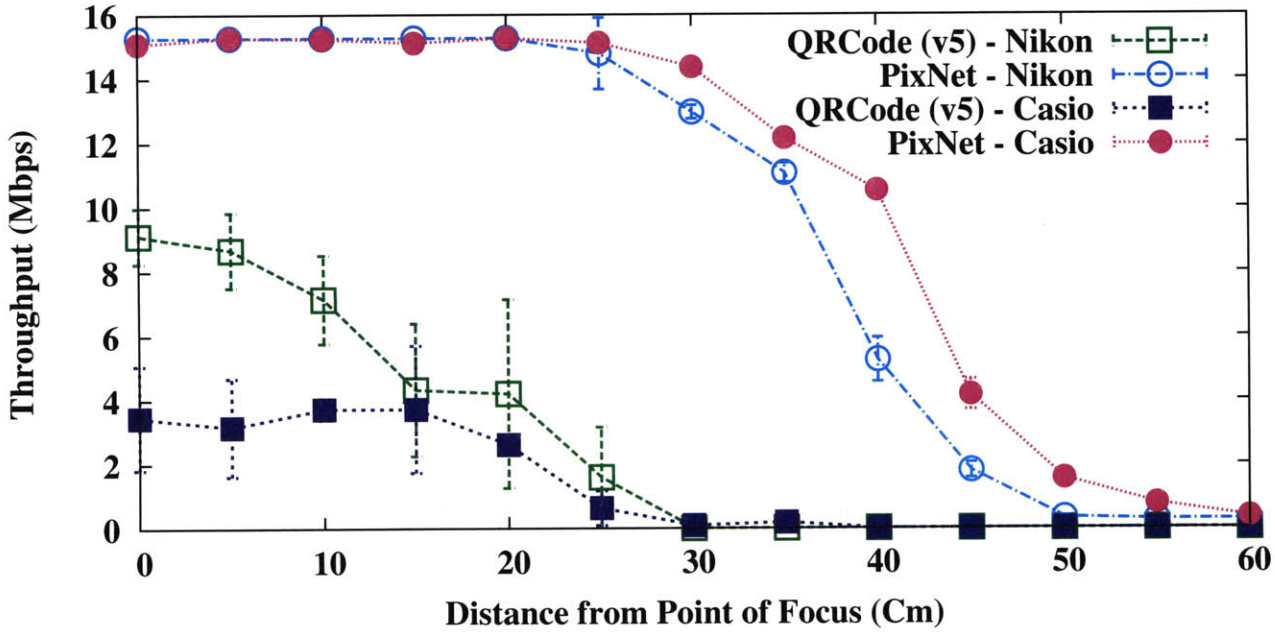


Figure 9-6—Impact of Blur on throughput achieved, using PixNet and QR Code. We find that PixNet is highly resiliant to blur and continues to perform reliably even while focussing at 40cm away from the plane of focus.

9.3 Impact of Viewing Angle

We evaluate PixNet and QR code with perspective distortion.

Method. In this experiment, the distance between transmitter and receiver is maintained at 2m and the view angle is changed to span the range $[0°, 140°]$ (view angles are measured as solid angles, where 140° refers to $-70°$ to 70°).

Results. Fig. 9-5 shows the throughput as a function of the view angle, both for PixNet and QR Code. PixNet is able to provide up to half its peak throughput (8 Mbps) at angles as wide as 120°, corresponding to a $3x$ gain in view angles compared with QR Code.

We note that while PixNet achieved longer distances with Casio, it achieves wider angles with Nikon. The reason for this behavior is that at very high angles, the transmitted rectangular image becomes a trapezoid at the receiver. The image size shrinks heavily on one side. As a result, only a few camera pixels are available to sample that portion of the image. Since Nikon has smaller pixels and a higher resolution than Casio, it can sample much more finely and retain more information. The performance of Casio is once again inferior for QR codes because of higher blur and coarser sampling.

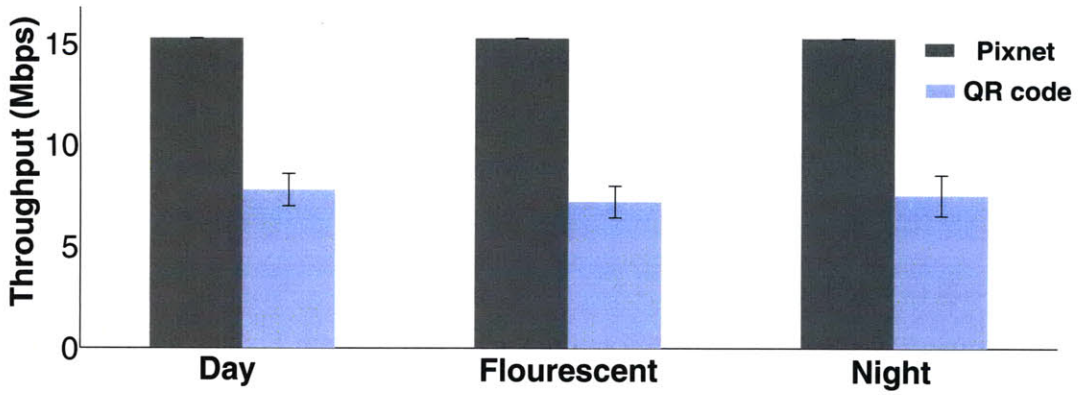


Figure 9-7—Impact of ambient light on throughput achieved, using PixNet and QR Code. We find that PixNet's performance is invariable with respect to ambient light. This stability in performance comes for free as an added advantage of operating in the frequency domain.

9.4 Impact of Blur

Blur is a common distortion in LCD-camera channels. It occurs because of focus imperfection, jerks while capturing images, or low resolution at the camera [24].

Method. To measure the throughput achieved as a function of the level of blur in the image, we use the same setup as in the previous experiments. We ensure that the frame displayed on the LCD is in sharp focus when the camera is at a distance of 2m from the LCD. We then progressively move the camera back away from the plane of focus. We take measurements at increasing distances without re-focusing the camera before capturing frames.

Results. Fig. 9-6 shows how PixNet's throughput changes with the amount of blur. PixNet provides constant throughput for distances up to 40 cm away from the plane of focus. On the other hand, QR code is highly sensitive to blur and rapidly deteriorates in performance even with a little amount of blur. More importantly, this result shows that PixNet does not degrade in performance even if the image experiences a reasonable amount of blur either due to handshakes or because it is not perfectly in focus.

We once again find that Casio performs better than Nikon for PixNet and performs worse than Nikon for QR code. The reason for this behavior is the difference in pixel sizes for Nikon and Casio as discussed before.

9.5 Impact of Ambient Light

For an LCD-camera link to provide high throughput in a variety of settings, its performance must remain stable over varying degrees of illumination.

Method. We fix the camera at a distance of 2m from the LCD. We keep the LCD and camera planes at zero view angle. We experiment with this setup for both PixNet and QR Code, under natural light, fluorescent light and in a dark room.

Result. In Figure 9-7, we present the throughput achieved by both PixNet and QR code. We observe that PixNet's performance is independent of the ambient light in the environment.

We observe that QR code's performance is also stable across different lighting conditions. QR code achieves this by employing a histogram based light-balancing technique that calculates an optimal intensity threshold in order to distinguish black pixels from white pixels. QR code's technique involves binning different intensity values and finding successive peaks and a valley in an intensity histogram. PixNet's scheme for light balancing however, is much simpler and requires no additional computation. It comes as an added advantage of operating in the frequency domain. Since illumination changes only effect the DC frequency, PixNet does not transmit any information in the DC frequency and simply discards the DC frequency at the receiver to prevent the variation in illumination from effecting the transmitted data.

9.6 Smart Phones

Smart phone cameras have smaller apertures than standard cameras [24]. This implies that the amount of light reaching a smart phone camera's sensors is typically an order of magnitude low (the intensity of light varies as a square of the aperture's diameter). And hence, the picture quality is usually poorer and has a lot more noise. Here, we would like to see what performance can be achieved by smart phone cameras.

Method. We use the setup in Fig. 9-1 with no appreciable angle between the camera and the LCD. We use Nokia N82 as our receiver. The LCD and the smart phone are separated by a distance of 2m.

Results. In Table 9.1, we show the throughput achieved for Nokia N82 for PixNet and QR code. We present results for Nikon and Casio receivers as well for comparison. Note that due to increased noise in the images captured by N82, QR codes yield no throughput at a granularity of 2. However, N82 is able to realize the maximum possible throughput that can be achieved with QR codes at a granularity of 5. We see

Camera Model	PixNet	QR code v5, g2	QR code v5, g5
Nokia N82	4.301 Mb/s	NILL	1.544 Mb/s
Nikon D3X	15.28 Mb/s	7.6673 Mb/s	1.55 Mb/s
Casio EX-F1	15.15 Mb/s	3.47 Mb/s	1.539 Mb/s

Table 9.1—Nominal Throughput in Mbps of PixNet and QR codes (version 5, granularities 2 and 5) across different receivers. PixNet's performance is superior even with smart phone cameras as receivers. Unlike QR codes, it requires no change in granularity or version number in order to function.

that PixNet performs well even for smart phones. Also note that since smart phone technology is rapidly evolving, we can expect to see smart phones equipped with higher quality cameras that will yield better performance.

9.7 Evaluation of 3D OFDM

In this section, we seek to compare the performance of PixNet with 3D OFDM to that of PixNet with 2D OFDM. Note that with 3D OFDM, the received frame rate is 60 fps and with 2D OFDM, the received frame rate is 30 fps. One might expect that since PixNet with 3D OFDM is able to operate at a higher rate, the throughput must be higher. However, it is also important to note that due to high attenuation in the higher frequencies, the higher frequencies are not used for transmission. And hence, the performance of PixNet with 3D OFDM when compared with that of PixNet with 2D OFDM can infact be inferior.

Method. We use the same setup as in Fig. 9-1 with no appreciable angle between the LCD and the camera. We use CASIO Ex-F1 as the receiver. The camera and the LCD are separted by a distance of 3m. We transmit a 2D encoded frame followed by a burst of 7 3D encoded frames.

Scheme	Throughput (mean)	Standard deviation
2D OFDM	11.05 Mbps	17.10 Kbps
3D OFDM	10.84 Mbps	1.92 Mbps

Table 9.2—Nominal Throughput of PixNet with 2D and 3D OFDM. We find that PixNet with 3D OFDM delivers the same throughput as PixNet with 2D OFDM.

Results. In Table 9.2, we show the throughput achieved for PixNet with 3D OFDM and PixNet with 2D OFDM. We find that the throughput of PixNet with 3D OFDM is in fact similar to the throughput achieved by PixNet with 2D OFDM. Though 3D OFDM carries the same amount of data while operating at twice the rate when compared with 2D OFDM, most of the high frequencies of the 3D OFDM system were unusuable because of inter-frame interference. And hence we see that 3D OFDM does not give us

as much gain as one might initially expect. We also find that there is a huge variance in the throughput achieved by 3D OFDM across different runs. This is because of the fact that the amount of inter-frame interference changes for every run depending on the timing offset between the camera's shutter and the display's graphics card (as described in §8.1).

However, another observation that's worth exploring is that the interference in the z direction (inter-frame interference) is limited only to the succeeding frame as opposed to the interference in the x or y directions which extends to more than a single neighboring pixel. Hence, coding along the z dimension can exploit this fact to design a potentially better 3D transmission scheme.

Chapter 10

Conclusion

We present PixNet, a novel approach to transmit data from an LCD to a camera, and show that such links can deliver high throughput (12-15 Mb/s) over multi-meter distances and wide view angles. LCD-camera pairs provide a new point in the design space of wireless communication. On the one hand, visible light signals cannot traverse walls and physical obstacles, thus limiting their communication range. On the other hand, LCD-camera links do not suffer from interference. We believe that spatial OFDM and the suite of tools introduced in this thesis offer a new perspective on the capability of these communication channels. LCD-camera links also enable an exciting array of applications ranging from using large LCD displays for broadcasting digital content to supporting point-to-point data transfer between handheld devices such as smart phones. We believe our work merely scratches the surface and a rich body of research lies ahead in this new and exciting field.

Appendix A

A mathematical model for phase offset computation

Proof of Proposition 5.2.2

Consider the set of four symbols forming a 2×2 super-symbol as shown above in Figure A-1. Without the loss of generality, we assume that there are only x corner offsets: A_x, B_x, C_x and D_x at corners C_1, C_2, C_3 and C_4 respectively. As we will see, it is straightforward to extend the analysis for the case when there are y corner offsets as well. Let L_s be the FFT length, L_g be the cyclic prefix length and $L_t = L_s + 2L_g$ be the total symbol length (the cyclic prefix is appended on both sides of the symbol). Let $W = 2L_s + 4L_g$ be the total width of the super-symbol. Let $s_{k,l,r}$ be the samples transmitted in the (k,l) frequency for symbol $r \in \{1,\ldots,4\}$. The 2D-OFDM symbols $S_{m,n,r}$ are formed by calculating the 2D-IFFT of $s_{k,l,r}$ as follows:

$$S_{m,n,r} = \sum_{m=0}^{L_s-1} \sum_{n=0}^{L_s-1} s_{k,l,r}\, e^{j\frac{2\pi km}{L_s}}\, e^{j\frac{2\pi ln}{L_s}}$$

The transmitted super-symbol R_{m_x,n_y} is formed by arranging the four symbols along with their cyclic prefix as shown in Fig. A-1. Now, to develop an equivalent continuous time model $R(x,y)$, let us assume that the samples at the transmitter, R_{m_x,n_y} are equally spaced with a time-period T such that

$$R_{m_x,n_y} = R(x,y)|_{x=m_xT,y=n_yT}$$

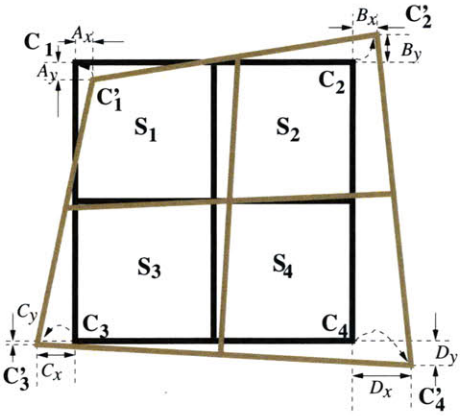


Figure A-1—A 2×2 super-symbol with x corner offsets of A_x, B_x, C_x and D_x at corners C_1, C_2, C_3 and C_4 respectively.

It then follows that

$$R(x,y) = \begin{cases} \displaystyle\sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1} s_{k,l,1}\, e^{j\frac{2\pi k(x-L_gT)}{L_sT}}\, e^{j\frac{2\pi l(y-L_gT)}{L_sT}} \\ \qquad \text{for } L_gT \le x < (L_s+L_g)T, \\ \qquad\qquad L_gT \le y < (L_s+L_g)T \\ \displaystyle\sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1} s_{k,l,2}\, e^{j\frac{2\pi k(x-(L_s+3L_gT))}{L_sT}}\, e^{j\frac{2\pi l(y-L_gT)}{L_sT}} \\ \qquad \text{for } (L_s+3L_g)T \le x < (2L_s+3L_g)T, \\ \qquad\qquad L_gT \le y < (L_s+L_g)T \\ \displaystyle\sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1} s_{k,l,3}\, e^{j\frac{2\pi k(x-L_gT)}{L_sT}}\, e^{j\frac{2\pi l(y-(L_s+3L_g)T)}{L_sT}} \\ \qquad \text{for } L_gT \le x < (L_s+L_g)T, \\ \qquad\qquad (L_s+3L_g)T \le y < (2L_s+3L_g)T \\ \displaystyle\sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1} s_{k,l,4}\, e^{j\frac{2\pi k(x-(L_s+3L_g)T)}{L_sT}}\, e^{j\frac{2\pi l(y-(L_s+3L_g)T)}{L_sT}} \\ \qquad \text{for } (L_s+3L_g)T \le x < (2L_s+3L_g)T, \\ \qquad\qquad (L_s+3L_g)T \le y < (2L_s+3L_g)T \end{cases}$$

Let $S'_{m,n,r}$ be the samples at the receiver obtained by sampling the super-symbol $R(x,y)$. If all the corners were detected accurately, then there would be no corner offsets and $S'_{m,n,r}$ would simply be $S_{m,n,r}$. However,

since we now have arbitrary x corner offsets, $S'_{m,n,r}$ is given by

$$S'_{m,n,1} = R\big((A_x + \tfrac{n_y}{W}(C_x - A_x))T + m_x T'_x,\ n_y T'_y\big)$$
$$S'_{m,n,2} = R\big((A_x + \tfrac{n_y}{W}(C_x - A_x))T + m_x T'_x,\ n_y T'_y\big)$$
$$S'_{m,n,3} = R\big((A_x + \tfrac{n_y}{W}(C_x - A_x))T + m_x T'_x,\ n_y T'_y\big)$$
$$S'_{m,n,4} = R\big((A_x + \tfrac{n_y}{W}(C_x - A_x))T + m_x T'_x,\ n_y T'_y\big)$$

where $T'_x = T(1 + \frac{(B_x - A_x)}{W} + n_y \frac{((D_x - C_x) - (B_x - A_x))}{W^2})$ and $T'_y = T$ (since there are no y corner offsets). $m_x = (m + L_g)$ for symbols 1, 3 and $(m + L_s + 3L_g)$ for symbols 2, 4. Similarly $n_y = (n + L_g)$ for symbols 1, 2 and $(n + L_s + 3L_g)$ for symbols 3 and 4. Also note that $\frac{|A_x| + |B_x| + |C_x| + |D_x|}{W} \sim \frac{1}{10W}$ since the corner detection scheme we use has an inaccuracy of only a few pixels over tens of super-symbols. Let us now consider symbol 1.

$$S'_{m,n,1} = \sum_{k=0}^{L_s-1} \sum_{l=0}^{L_s-1} s_{k,l,1}\, e^{j\frac{2\pi k(mT'_x + L_g(T'_x - T))}{L_s T}}\, e^{j\frac{2\pi l(nT'_y + L_g(T'_y - T))}{L_s T}} \cdots e^{j\frac{2\pi k(A_x + (C_x - A_x)\frac{n+L_g}{W})}{L_s}}$$
$$= \sum_{k=0}^{L_s-1} \sum_{l=0}^{L_s-1} s_{k,l,1}\, e^{j\frac{2\pi km}{L_s}}\, e^{j\frac{2\pi ln}{L_s}} \cdots e^{j\frac{2\pi k}{L_s}\left(A_x + \frac{((n+L_g)(C_x - A_x) + (m+L_g)(B_x - A_x))}{W} + \frac{(m+L_g)(n+L_g)((D_x - C_x) - (B_x - A_x))}{W^2}\right)}$$

The transmitted samples $s_{k,l,1}$ are recovered at the receiver as $z_{p,q,1}$ by taking the 2D FFT of the received

samples $S'_{m,n,1}$

$$z_{p,q,1} = \frac{1}{L_s^2} \sum_{m=0}^{L_s-1} \sum_{n=0}^{L_s-1} S'_{m,n,1}\, e^{-j\frac{2\pi mp}{L_s}}\, e^{-j\frac{2\pi nq}{L_s}}$$

$$= \frac{1}{L_s^2} \sum_{\substack{0\le k<L_s \\ 0\le l<L_s}} s_{k,l,1} \sum_{\substack{0\le m<L_s \\ 0\le n<L_s}} e^{j\frac{2\pi m(k-p)}{L_s}}\, e^{j\frac{2\pi n(l-q)}{L_s}} \ \dots$$

$$e^{j\frac{2\pi k}{L_s}\left(A_x+\frac{((n+L_g)(C_x-A_x)+(m+L_g)(B_x-A_x))}{W}+\frac{(m+L_g)(n+L_g)((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$

$$= \frac{1}{L_s^2} \sum_{\substack{0\le k<L_s \\ 0\le l<L_s}} s_{k,l,1}\, e^{j\frac{2\pi k}{L_s}\left(A_x+\frac{L_g((C_x-A_x)+(B_x-A_x))}{W}+\frac{L_g^2((D_x-C_x)-(B_x-A_x))}{W^2}\right)} \dots$$

$$\sum_{m=0}^{L_s-1} e^{j\frac{2\pi m(k-p)}{L_s}}\, e^{j\frac{2\pi km}{L_s}\left(\frac{(B_x-A_x)}{W}+\frac{L_g((D_x-C_x)-(B_x-A_x))}{W^2}\right)} \dots$$

$$\sum_{n=0}^{L_s-1} e^{j\frac{2\pi n}{L_s}\left((l-q)+\frac{k(C_x-A_x)}{W}+\frac{k(m+L_g)((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$

Applying Lemma 1 (below) for the summation over n, we have $z_{r,s,1}$

$$\simeq \frac{1}{L_s} \sum_{\substack{0\le k<L_s \\ 0\le l<L_s}} s_{k,l,1}\, e^{j\frac{2\pi k}{L_s}\left(A_x+\frac{L_g((C_x-A_x)+(B_x-A_x))}{W}+\frac{L_g^2((D_x-C_x)-(B_x-A_x))}{W^2}\right)} \dots$$

$$\sum_{m=0}^{L_s-1} e^{j\frac{2\pi m(k-p)}{L_s}}\, e^{j\frac{2\pi km}{L_s}\left(\frac{(B_x-A_x)}{W}+\frac{L_g((D_x-C_x)-(B_x-A_x))}{W^2}\right)} \dots$$

$$\delta(l-q)\, e^{j\pi\left(\frac{k(C_x-A_x)}{W}+\frac{k(m+L_g)((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$

$$\simeq \frac{1}{L_s} \sum_{\substack{0\le k<L_s \\ 0\le l<L_s}} s_{k,l,1}\delta(l-q) \dots$$

$$e^{j\frac{2\pi k}{L_s}\left(A_x+\frac{(L_g+\frac{L_s}{2})(C_x-A_x)+L_g(B_x-A_x))}{W}+\frac{L_g(L_g+\frac{L_s}{2})((D_x-C_x)-(B_x-A_x))}{W^2}\right)} \dots$$

$$\sum_{m=0}^{L_s-1} e^{j\frac{2\pi m}{L_s}\left((k-p)+\frac{k(B_x-A_x)}{W}+\frac{k(L_g+\frac{L_s}{2})((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$

Applying Lemma 1 for the summation over m, we have $z_{r,s,1}$

$$\simeq \sum_{\substack{0\le k<L_s \\ 0\le l<L_s}} s_{k,l,1}\delta(l-q)\delta(k-p)\ldots$$
$$e^{j\frac{2\pi k}{L_s}\left(A_x+\frac{(L_g+\frac{L_s}{2})((C_x-A_x)+(B_x-A_x))}{W}+\frac{(L_g+\frac{L_s}{2})^2((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$
$$\simeq s_{p,q,1}e^{j\frac{2\pi p}{L_s}\left(A_x+\frac{(L_g+\frac{L_s}{2})((C_x-A_x)+(B_x-A_x))}{W}+\frac{(L_g+\frac{L_s}{2})^2((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$
$$\simeq s_{p,q,1}\,e^{j\frac{2\pi p}{16L_s}(9A_x+3B_x+3C_x+D_x)} \qquad (\text{Since } W=4L_g+2L_s)$$

Similarly for symbol 2,

$$S'_{m,n,2}=\sum_{k=0}^{L_s-1}\sum_{l=0}^{L_s-1}s_{k,l,2}\,e^{j\frac{2\pi km}{L_s}}\,e^{j\frac{2\pi ln}{L_s}}\ \ldots$$
$$e^{j\frac{2\pi k}{L_s}\left(A_x+\frac{((n+L_g)(C_x-A_x)+(m+L_s+3L_g)(B_x-A_x))}{W}+\frac{(m+L_s+3L_g)(n+L_g)((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$

And hence it follows that

$$z_{p,q,2}\simeq s_{p,q,2}\,e^{j\frac{2\pi p}{L_s}\left(A_x+\frac{((L_g+\frac{L_s}{2})(C_x-A_x)+(\frac{3L_s}{2}+3L_g)(B_x-A_x))}{W}\right)}\ldots$$
$$e^{j\frac{2\pi p}{L_s}\left(\frac{(\frac{3L_s}{2}+3L_g)(\frac{L_s}{2}+L_g)((D_x-C_x)-(B_x-A_x))}{W^2}\right)}$$
$$\simeq s_{p,q,2}\,e^{j\frac{2\pi p}{16L_s}(3A_x+9B_x+C_x+3D_x)}$$

Likewise, $z_{p,q,3}\simeq s_{p,q,3}\,e^{j\frac{2\pi p}{16L_s}(3A_x+B_x+9C_x+3D_x)}$

$$z_{p,q,4}\simeq s_{p,q,4}\,e^{j\frac{2\pi p}{16L_s}(A_x+3B_x+3C_x+9D_x)}$$

Therefore, we have the following relationship between the corner offsets A_x, B_x, C_x and D_x and the resulting differences in phase shifts $\Delta\theta_{r,x}$, $r\in\{1,\ldots,4\}$

$$\begin{pmatrix}\Delta\theta_{1,x}\\ \Delta\theta_{2,x}\\ \Delta\theta_{3,x}\\ \Delta\theta_{4,x}\end{pmatrix}\simeq\frac{2\pi(k-k')}{16L_s}\begin{pmatrix}9&3&3&1\\3&9&1&3\\3&1&9&3\\1&3&3&9\end{pmatrix}\begin{pmatrix}A_x\\B_x\\C_x\\D_x\end{pmatrix}$$

Note that a corner offset along the x direction results in phase offsets only in the x frequencies at the receiver. It is straightforward to derive a similar relationship between y corner offsets and the differences in phase offsets in the y frequencies.

Lemma 1: Sum of complex exponents

$$\sum_{n=0}^{L_s-1} e^{j\frac{2\pi n}{L_s}(p+\xi)} \simeq L_s\, e^{j\pi\xi}\delta(p)$$

$$\text{for } |\xi| < \tfrac{1}{4\pi} \text{ and } 0 \le |p| < L_s, p \in \mathbb{Z}.$$

Proof of Lemma:

$$\sum_{n=0}^{L_s-1} e^{j\frac{2\pi n}{L_s}(p+\xi)} = \frac{1-e^{j2\pi(p+\xi)}}{1-e^{j\frac{2\pi(p+\xi)}{L_s}}} \qquad \left(\text{Since } \sum_{n=0}^{L_s-1} x^n = \frac{1-x^{L_s}}{1-x}\right)$$

$$= \frac{e^{j\pi(p+\xi)}}{e^{j\frac{(p+\xi)}{L_s}}} \frac{\sin\big(\pi(p+\xi)\big)}{\sin\big(\frac{\pi(p+\xi)}{L_s}\big)}$$

$$\simeq e^{j\pi(p+\xi)} \frac{\sin\big(\pi(p+\xi)\big)}{\sin\big(\frac{\pi(p+\xi)}{L_s}\big)} \qquad (\text{Since } L_s >> 1)$$

case 1: $p = 0$

$$\frac{\sin(\pi\xi)}{\sin\big(\frac{\pi\xi}{L_s}\big)} \simeq \frac{\pi\xi}{\frac{\pi\xi}{L_s}} = L_s \qquad (\text{Since } \pi|\xi| < \tfrac{1}{4})$$

case 2: $p \neq 0$

$$\left|\frac{\sin\big(\pi(p+\xi)\big)}{\sin\big(\frac{\pi(p+\xi)}{L_s}\big)}\right| = \left|\frac{\sin(\pi\xi)}{\sin\big(\frac{\pi(p+\xi)}{L_s}\big)}\right| < \frac{|\sin(\pi\xi)|}{\min\limits_{p\neq 0}\left|\sin\big(\frac{\pi(p+\xi)}{L_s}\big)\right|}$$

Now $\left|\sin\big(\frac{\pi(p+\xi)}{L_s}\big)\right|$ is the smallest when $p = 0$. However since $p \neq 0$ and $\sin(x) \simeq 0$ when $x \simeq 0$ or π, the next best value is $p = \pm 1$ or $\pm(L_s - 1)$. And hence, $\min\limits_{p\neq 0}\left|\sin\big(\frac{\pi(p+\xi)}{L_s}\big)\right| = \sin\big(\frac{\pi(1-|\xi|)}{L_s}\big)$. Therefore we have,

$$\left|\frac{\sin\big(\pi(p+\xi)\big)}{\sin\big(\frac{\pi(p+\xi)}{L_s}\big)}\right| < \frac{\big|\sin(\pi\xi)\big|}{\big|\sin\big(\frac{\pi(1-|\xi|)}{L_s}\big)\big|} \qquad (\text{Since } p \neq 0)$$

$$\simeq \frac{\pi|\xi|}{\frac{\pi(1-|\xi|)}{L_s}} \qquad (\text{Since } \pi|\xi| < \tfrac{1}{4};\ L_s >> 1)$$

$$= \frac{L_s}{(\frac{1}{|\xi|}-1)} < \frac{L_s}{(4\pi-1)} << L_s$$

And hence,

$$\sum_{n=0}^{L_s-1} e^{j\frac{2\pi n}{L_s}(p+\xi)} \simeq e^{j\pi(p+\xi)}\frac{\sin\big(\pi(p+\xi)\big)}{\sin\big(\frac{\pi(p+\xi)}{L_s}\big)}$$

$$\simeq L_s\, e^{j\pi\xi}\delta(p)$$

Bibliography

[1] Automatic identification and data capture techniques – Aztec code bar code symbology specification. ISO/IEC 24778:2008.

[2] Automatic identification and data capture techniques – QR code 2005 bar code symbology specification. ISO/IEC 18004:2006.

[3] Automatic identification and data capture techniques - Data Matrix bar code symbology specification. ISO/IEC 16022:2006.

[4] International symbology specification – maxicode. ISO/IEC 16023:2000.

[5] *Digital Image Warping*. IEEE computer society Press, 1990.

[6] *Reed-Solomon Codes and Their Applications*. Wiley-IEEE Pres, 1999.

[7] *Multi Carrier Digital Communications: Theory and Applications of OFDM*. Springer New York, 2004.

[8] Shilpi Gupta and Christopher Jaynes. Active pursuit tracking in a projector-camera system with application to augmented reality. In *CVPR*, 2005.

[9] P. R. Haugen, Rychnovsky S, Husain A, and Hutchenson L. D. Optical interconnects for high speed computing. *OPT. ENG.*, 25(10):1076–1085, 1986.

[10] HCCB. http://research.microsoft.com/en-us/projects/hccb/.

[11] S. Hranilovic and F. R. Kschischang. A pixelated mimo wireless optical communication system. In *IEEE Journal of Selected Topics in Quantum Electronics*, 2006.

[12] H. Kato and K.T. Tan. 2d barcodes for mobile phones. In *2nd Int'l Conf. Mobile Technology, Applications and Systems (MTAS 2005)*, pages P1A–4.

[13] C.G. Lee, C. S. Park, J. H. Kim, and D. H. Kim. Experimental verification of optical wireless communication link using high-brightness illumination light-emitting diodes. *Fiber Optics and Optical Comm.*, Dec 2007.

[14] Hana Lee and Jeongtae Kim. Retrospective correction of nonuniform illumination on bi-level images. *Opt. Express*, 17, 2009.

[15] David A.B. Miller. Optical interconnects to silicon. *IEEE Journal on selected topics in quantum electronics*, 6(6), 2000.

[16] Hoa Le Minh, Dominic O'Brien, Grahame Faulkner, Lubin Zeng, Kyungwoo Lee, Daekwang Jung, and YunJe Oh. High-speed visible light communications using multiple-resonant equalization. In *IEEE Photonics*, pages 1243–1245, 2008.

[17] Ankit Mohan, Grace Woo, Shinsaku Hiura, Quinn Smithwick, and Ramesh Raskar. Bokode: imperceptible visual tags for camera based interaction from a distance. In *ACM SIGGRAPH*, 2009.

[18] Dominic O'Brien, Hoa Le Minh, Lubin Zeng, Grahame Faulkner, Kyungwoo Lee, Daekwang Jung, YunJe Oh, and Eun Tae Won. Indoor visible light communications: challenges and prospects. In *SPIE Optics and Photonics*, volume 7091, pages 709106–709106–9, 2008.

[19] Eisaku Ohbuchi, Hiroshi Hanaizumi, and Lim Ah Hock. Barcode readers using the camera device in mobile phones. In *International Conference on Cyberworlds (CW'04)*, 2004.

[20] openGL. http://www.opengl.org.

[21] J. R. Parker. *Algorithms for Image Processing and Computer Vision*. Wiley, 1996.

[22] C. Pinhanez. Using a steerable projector and camera to transform surfaces into interactive displays, 2001.

[23] Morton A. Q. Packaging history: The emergence of the uniform product code (upc) in the united states. *History and Technology*, 11(1):101–111, 1994.

[24] Ramesh Raskar and Jack Tumblin. *Computational Photography: Mastering New Techniques for Lenses, Lighting, and Sensors*. A K Peters, 2010.

[25] J. Rekimoto and Y. Ayatsuka. Cybercode: Designing augmented reality environments with visual tags. In *DARE*. ACM Press, 2000.

[26] Michael Rohs. Real-world interaction with camera phones. *Lecture Notes in Computer Science*, 2005.

[27] Grace Woo, Ankit Mohan, Ramesh Raskar, and Dina Katabi. Simple LCD transmitter camera receiver data link. Technical report, CSAIL, Massachusetts Institute of Technology, 2009. http://hdl.handle.net/1721.1/4556.

[28] Zxing. http://code.google.com/p/zxing/.