國立臺灣大學電機資訊學院資訊工程學系
碩士論文
Department of Computer Science and Information Engineering
College of Electrical Engineering and Computer Science
National Taiwan University
Master Thesis

# 現實環境中螢幕-相機通訊系統之實驗分析

# Experimental Analysis of Real-World Screen-to-Camera Communications

張乃中
Nai-Chung Chang

指導教授：蔡欣穆博士
Advisor: Hsin-Mu Tsai, Ph.D.

中華民國 108 年 7 月
July, 2019

doi:10.6342/NTU201903432

# 國立臺灣大學碩士學位論文
# 口試委員會審定書

## 現實環境中螢幕-相機通訊系統之實驗分析

## Experimental Analysis of Real-World Screen-to -Camera Communications

本論文係張乃中君（學號 R05922112）在國立臺灣大學資訊工程學系完成之碩士學位論文，於民國 108 年 7 月 29 日承下列考試委員審查通過及口試及格，特此證明

口試委員：蔡欣穆

（指導教授）

施吉昇　　蘇志宇

黃琴雅

系主任　莊永裕

# 誌謝

本篇論文能夠完成要感謝我的指導教授蔡欣穆老師，從研究題目的挑選、實驗方法的討論與結果內容的分析都給予我不少的協助與指導，在研究遇到瓶頸時也會適時提點我實驗中存在的問題。老師也讓我了解到處事態度與事物細節的重要性，在日常生活中也會給予我人生經歷與興趣上的建議與鼓勵。

此外要特別感謝的是雯萱，在忙碌之中還是與我討論研究內容、逐字修改我的論文文句且不厭其煩的協助我準備學位口試，幫助我完成一本正式的碩士論文。且在我灰心喪志時給予我不少勉勵，在閒餘之時又是個能互相分享生活瑣事的朋友。此外實驗室的各位都給予我碩士學涯中各種的協助：Wally、孟庭和家豐於研究上的討論；豆豆、才智與浩平於生活中的分享；楷楷於日常中的關心而韋辰於視野上的開拓，及實驗室各位給予我的回憶。

另外要感謝我的朋友兼室友宗德，總是一針見血的給予我建議，聆聽我的怨言及忍受我的喧鬧；高中學長小花的勉勵以及大學朋友們的陪伴，舒緩我碩士學涯的鬱悶，在我支撐不住時給了我堅持下去的動力，尤其是不斷給予我鼓勵的佳勳和陪我品嚐台北的品睿。最後要感謝在背後默默支持我的家人，有你們才能讓我在無後顧之憂的情況下完成碩士學位，並在有限的人生中讓我自由且任性地去見識這個世界、懂得如何去解讀和關心這個社會，而不是成為那些冷漠之人。但願本篇論文能對正在翻閱的你有所幫助。



doi:10.6342/NTU201903432

# 摘要

螢幕-相機通訊系統為現今最為常見的一種短距離通訊系統，空間域編碼方式的螢幕-相機通訊系統目前已被廣泛使用於生活當中。然而使用空間域編碼方式的螢幕-相機通訊系統也面臨了一些挑戰，包括通訊距離的不足、容易受到影像模糊的影響和資料傳輸量上的限制，導致空間域編碼方式的螢幕-相機通訊系統無法適用於更多的應用場景。為了改善螢幕-相機通訊系統之通訊距離不足問題，我們選擇較不容易受到影像模糊影響之頻率域編碼方式來設計傳輸端呈現的二維碼。本篇論文中三個主要的挑戰為螢幕-相機間亮度的非線性關係、二維碼的透視變形和影像模糊的問題。針對這三個挑戰，我們分別提出亮度校正、重取樣與通道估測的方法來降低解碼的錯誤率。為探討這些方法的效果，我們採用三種大小不同的螢幕及兩種相機來進行測試。實驗結果顯示，亮度校正能有效的降低至多 20% 的通訊錯誤率。而重採樣的方法則能有效的降低圖形形變造成的解碼錯誤。

關鍵字: 螢幕-相機通訊系統、可見光通訊、頻率域編碼方式



doi:10.6342/NTU201903432

# Abstract

Screen-to-camera communication is one of the most popular short-range communication systems in recent years.Two main encoding schemes used in screen-to-camera communications are spatial encoding and frequency encoding scheme. Even though spatial encoding scheme, such as QR code, has been widely deployed in our daily life, spatial encoding scheme suffers from short communication range, and is vulnerable to the blurring effect. On the other hand, frequency encoding scheme is considered to be more resilient to the blurring effect. As our ultimate goal is to increase the communication range of screen-to-camera communications, frequency encoding scheme is adopted in this work to generate the transmitted code. There are three major challenges in this work: nonlinear relationship between the transmitted intensity by the screen and received intensity by the camera, perspective distortion and the blurring effect. To address these challenges, we proposed nonlinearity calibration, resampling and channel estimation respectively to reduce the decoding error. To evaluate our proposed methods, three kinds of screens with different sizes and two cameras are used to carry out the experiments. The experimental results show that nonlinearity calibration is able to reduce the error rate with an error rate drop of at most 20%. In addition, the resampling algorithm can effectively mitigate the error caused by the distortion.

Keyword: Screen-to-camera communication, Visible light communication, Frequency encoding scheme



doi:10.6342/NTU201903432

# Contents





doi:10.6342/NTU201903432

doi:10.6342/NTU201903432

# List of Figures





doi:10.6342/NTU201903432

# List of Tables





doi:10.6342/NTU201903432

# Chapter 1

# Introduction

Due to the rapid development of technology, mobile devices such as mobile phones or tablets are ubiquitous in our daily life. Cameras, as one of the essential components in these mobile devices, have become more and more common. it offers a great opportunity to develop a unique wireless communication paradigm called screen-to-camera communication. Screen-to-camera communications have emerged as one of the most popular technologies for providing one-way data transmission in a short range. Information is encoded into a transmitted code which is shown on a monitor or a screen-equipped device as the transmitter. Whereas, the most common receiver used for capturing transmitted codes are the commodity cameras in mobile phones. Compare to the wireless radio frequency technology such as Bluetooth, which is also applied in short-range communication, screen-to-camera communications provides a more convenient way to transmit data without any synchronization or pairing process between the transmitter and receiver.

In prior works, the spatial encoding method is the most widely used method for turning the transmitted information into images (i.e., transmitted codes). The transmitted information is first translated into binary datas which can be then expressed as block cells with color of white and black, forming a transmitted code. One of the most common examples of spatial encoding method is QR code, which stands for Quick Response Code. QR code enables a simple and convenient scenario for data exchange. It usually appears on advertisements and websites to transmit the information to the public. Moreover, QR code can be used as the major medium in mobile payment and communication software for



doi:10.6342/NTU201903432

exchanging personal information.

Despite the fact that the design of the QR code continuously evolves to meet the needs of users, in its current form it still presents challenges in long-distance scenarios. When people want to obtain the information from a QR code, it is crucial to put the camera as close to the QR code as possible to make the code in the captured image large enough for recognition and decoding. As a result, the transmission distance of the QR code is usually less than one meter [12], which is problematic for many applications. For example, the museum usually uses QR code to provide more information about the artwork to visitors. Due to the aesthetic perception, QR code usually appears in a small form, and would have a difficult time capturing the code when it is crowded. The main reason that makes QR code unable to work at a long-range distance is the blurring effect, which is caused by the insufficient sharpness of the camera lens or the slightly shaking when capture the image. The blurring effect would cause the color of adjacent block cells to be blended with each other, leading to false decode of the transmitted QR code.

To decrease the influence of the blurring effect, a new scheme which encodes the data in the frequency domain instead of the spatial domain had been proposed in recent years, providing a new perspective for implementing the screen-to-camera communication system [6]. This is due to the fact that the blurring effect caused by the optical system in the camera produce similar effect to the low-pass filter, causing the attenuation at high frequency while the low frequency components remain unchanged. Therefore, frequency-based screen-to-camera communication features an innate resistance to the blurring effect which naturally reduces the the range of affected frequencies by blurring effect and minimize the number of bit errors.

PixNet [11] is the first work for implementing the frequency-based screen-to-camera communication system. Based on [6], PixNet generalizes the OFDM scheme to address the issue of perspective distortion. However, PixNet did not address the issue of blurring effect in a fundamental manner. PixNet disregards the channel at high frequency and does not make use of it, while protecting other frequencies with a high redundancy error correction code. Focus [5], which builds on the concept from OFDM, partitions the frequency



doi:10.6342/NTU201903432

spectrum to multiple sub-channels. The system encodes different amount of data, according to the robustness of the sub-channel. To lower the error rate caused by the frame mixing, it presents multi-rate streams that enable a transmitter to concurrently support multiple receivers (i.e., cameras) with different frame rates. Unfortunately, the proposed system focuses on the design of code construction and the issue of frame mixing. The impact of blurring effect has not been mitigated properly. To provide better resilience in the screen-to-camera communication system in various scenarios, we proposed several methods for improving the working range and the adaptability of the screen-to-camera communication system.

In this work, we take advantage of the frequency encoding scheme and focus on improving the working range of screen-to-camera communication while reducing the erroneous data caused by distance during the transmission. Due to the increasing transmission distance, we need to address three major challenges.

First, the nonlinearity between the transmitted and received intensity values produces large decoding errors, degrading system performance. This is because that the brightness displayed on the screen is obviously different from the intensity received by the camera, depending heavily on the exposure setting of the camera and the capability of the image sensor. The nonlinearity between the transmitted and received intensity makes the received intensity no longer be proportional to the transmitted intensity, causing decoding errors. To address this issue, we design a gradient pattern to measure the relative relationship of the luminance between the transmitter and the receiver, and utilize the nonlinearity calibration information to mitigate the error.

Second, when the camera is not perfectly aligned with the screen, perspective distortion occurs due to non-perpendicular viewing angle between the transmitting screen and the receiving camera. Perspective distortion induces code deformation when the transmitted code projects on the image sensor. The deformation of the transmitted code leads to an inconsistent sampling frequency in the spatial domain and resulting in the frequency shift in the frequency domain. To deal with the perspective distortion, we apply the common method to eliminate the sampling frequency offset in the OFDM and generalize it to



doi:10.6342/NTU201903432

two-dimensional spatial signal.

Finally, the blurring effect is considered to be one of the main causes that degrades the performance of screen-to-camera communication. Blurring effect is the result of pixel blending together with adjacent pixels at the receiving camera, which is caused by insufficient sharpness of the lens and the defocus of the camera. The imperfectness of the lens structure causes the light to be slightly divergent when passing through the lens, resulting in the color-blending pixels which eliminate the details of the signal and cause information loss at high frequency. In the case of longer transmission distance, the transmitted code occupies fewer pixels in the captured image, where the blurring effect would become more severe. In addition, the imperfect alignment between the symbol blocks in the transmitted code and the pixels in the image sensor causes color-blending problem. For example, a symbol block in the transmitted code would be smaller than a pixel in the receiving camera at long distance. Therefore, a pixel in the image sensor would sense the intensity of multiple symbol blocks in the transmitted code. As the pixel would output the total intensity of its sensing area, the output intensity of the pixel would be the sum of all the intensity observed symbol blocks. To address this issue, we estimate the channel using pilot symbols in the frequency domain, and utilize the channel information to correct the received signal at high frequency.

This thesis provides an analysis on the issues of screen-to-camera communications which causes system degradation in long-range communications. To address these issues, we propose several methods to improve the performance of screen-to-camera communications, and report the following key findings:

- We adopt non-linearity correction in our design, which decreases the sensitivity difference of the luminance between the screen and camera. It reduces the error rate significantly in higher-order modulation.
- We improve the resampling method which can mitigate performance degradation due to perspective distortion. We address the blurring effect by estimating the blurring channel. With the estimated channel, the receiver can reconstruct the transmitted signal with more detail, mitigating the error.



doi:10.6342/NTU201903432

- We evaluate the improved screen-to-camera communications with the schemes in various situations with different devices. The results show that the improved screen-to-camera communication has better performance at a long distance.

The rest of this thesis is organized as follows. In Chapter 2, literature in both spatial-based and frequency-based screen-to-camera communication are described. Background knowledge of screen-to-camera communication concepts is mentioned in Chapter 3. Chapter 4 introduces the design of the whole system and the algorithms we used in our system. The implementation procedures are present in Chapter 5. The system evaluation and analysis result are shown in Chapter 6. Finally, Chapter 7 concludes this thesis.



doi:10.6342/NTU201903432

# Chapter 2

# Related Work

In screen-to-camera communications, spatial-based and frequency-based encoding schemes are the two most commonly used methods for encoding the transmitted data into a 2D image. While spatial-based encoding scheme directly uses the pixel blocks to represent the transmitted data in the spatial domain, frequency-based encoding scheme, however, utilizes the frequency channels to encode the data. In the following, we will introduce several prior works on spatial-based or frequency-based screen-to-camera communications.

## 2.1 Spatial-based screen-to-camera communication

Spatial-based screen-to-camera communication becomes popular for short-range communication in recent years. Various design of 2D barcode such as QR Code [3], Data Matrix [1] and Aztec codes [2] has been established for practical application. To achieve better performance and efficiency, most of the papers focused on improving the efficiency of the 2D barcode. Due to the progress of image processing, the deformation correction has been proposed for identifying the QR Code more accurately [13]. In order to meet the increasing demand of the transmitted data, the design of the 2D barcode has been improved for the higher data capacity [13, 10, 4]. Moreover, to extend the workable transmission scenarios, spatial-based screen-to-camera communication had been analyzed in different environments and attempted to increase the adaptivity of the 2D barcode [9, 8, **?**].

Among these papers, there are several papers aiming to increase the transmission dis-



doi:10.6342/NTU201903432

tance or mitigate the impact of the blurring effect. COBRA [4] developed a novel 2D color barcode on smartphones, which could achieve real-time decoding for barcode streaming. COBRA not only adopted a blur-aware color ordering to alleviate the image blur in mobile environments but also proposed an adaptive code generation scheme to adjust the block size adaptively based on the mobility of the smartphone. Strata [7], on the other hand, proposed a layered coding scheme for spatial-based screen-to-camera communication, which enables the transmitter to encode the data with multiple resolutions in the spatial domain. Depends on the transmission distance and capability of the camera, the data with different resolutions can be decoded. Although these papers all put an effort on reducing the influence of the blurring effect, there still exists a trade-off between the data capacity and the transmission distance.

## 2.2 Frequency-based screen-to-camera communication

Due to the rising demand for higher throughput and transmission distance, more efficient and stable screen-to-camera communication systems are expected . The concept of applying OFDM to screen-to-camera communication was first proposed by Hranilovic and Kschischang [6]. The idea was inspired by the multiple-input/multiple-output (MIMO) systems at radio frequencies (RF). Experimental prototype are carried out to validate feasibility of the idea.

PixNet [11], which built on the previous work [6], implemented the first system of frequency -based screen-to-camera communication. PixNet generalized the correction method of sampling frequency offset in OFDM to address the perspective distortion and applied the blur-adaptive error correction code to deal with the impact of blurring effect. The blur-adaptive coding scheme presented by PixNet discards the bandwidth in the high-frequency channels, while high redundancy error correction code is added the other frequency channels to ensure system robustness. Rather than throwing the high frequency away, Focus [5] partitions the frequency spectrum into many separate sub-channels. Each sub-channel encodes one part of the payload. The receiver is able to adaptively decode different amount of data according to the transmission distance and camera resolution.



doi:10.6342/NTU201903432

Focus also presents a multi-rate streams scheme which enables different receivers to use its own frame rate to capture the encoded data. However, None of the above works effectively tackle the issue of blurring effect. To this end, in this work, we propose to address the degradation resulting from the blurring effect by introducing nonlinearity calibration technique and channel estimation algorithm. In addition, we also improve the perspective correction algorithm introduced in PixNet to enhance the system resilience to various viewing angles between the transmitter and the receiver.



doi:10.6342/NTU201903432

# Chapter 3

# PRELIMINARY

## 3.1 2D Orthogonal Frequency Division Multiplexing

Similar to traditional RF signals which can be transformed from time domain to frequency domain, an image can be transformed from spatial domain to "spatial frequency" domain using 2D Fourier transform. Fig shows an example of the spatial frequency spectrum with size $N \times N$, where the frequency block in the center is considered the DC frequency representing for the average luminance of the entire image. The frequency blocks that are farther to the center stand for higher frequencies.

As a result, one can encode the transmitted data in the frequency domain and transform the data into spatial domain as an image. This encoding process is regarded as a variant of Orthogonal Frequency Division Multiplexing (OFDM), being called 2D OFDM in [PixNet]. To create a 2D OFDM symbol, 2D Inverse Fourier transform is applied to a $N \times N$ data matrix $A$, where $a_{i,j}$ is an element in $A$ which is a modulated data bit. The expression of a 2D OFDM symbol S is

$$S_{p,q} = \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} s_{k,l} e^{j\frac{2\pi kp}{N}} e^{j\frac{2\pi lq}{N}} \tag{3.1}$$

On the other hand, if the receiver samples the symbol accurately, the transmitted data



doi:10.6342/NTU201903432

of this OFDM symbol $a$ can be decoded using FFT:

$$a'_{p,q} = \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} S_{k,l} e^{-j\frac{2\pi kp}{N}} e^{-j\frac{2\pi lq}{N}} \tag{3.2}$$

## 3.2 RANSAC

Random sample consensus, also known as RANSAC, is an iterative method to estimate the parameter of a mathematical model from a set of observed data. It is a non-deterministic algorithm which is randomly select the observed data to establish the model. The advantage of RANSAC is that it has great resistance against the influence of outliers during the model estimation, which would produce a reasonable result according to the number of iterations.

In a set of observed data, the data could be usually classified into inliers and outliers. Inliers refer to the data which roughly could be fitted to the model, while outliers refer to the noise or the deviation in the data. In the beginning, RANSAC would randomly select several data in the set of observed data to calculate the model and uses the model to classify the data. After several selections, according to the probability, the model we establish would have a high possibility to fit most of the data, which is also approximate to the best solution. The algorithm applies by repeating the following steps:

1. Randomly select the required number of observed data to determine the model parameters.
2. Calculate for the parameters of the model.
3. Determine the number of the inliers, which is fit to the model with a predefined tolerance value from the rest of the observed data.
4. Store the number of inliers in the consensus set.
5. Repeat steps 1 through 4 (maximum of N times).
6. Select the model with the most inliers as the optimal solution



doi:10.6342/NTU201903432

# Chapter 4

# System Design

## 4.1 Overview

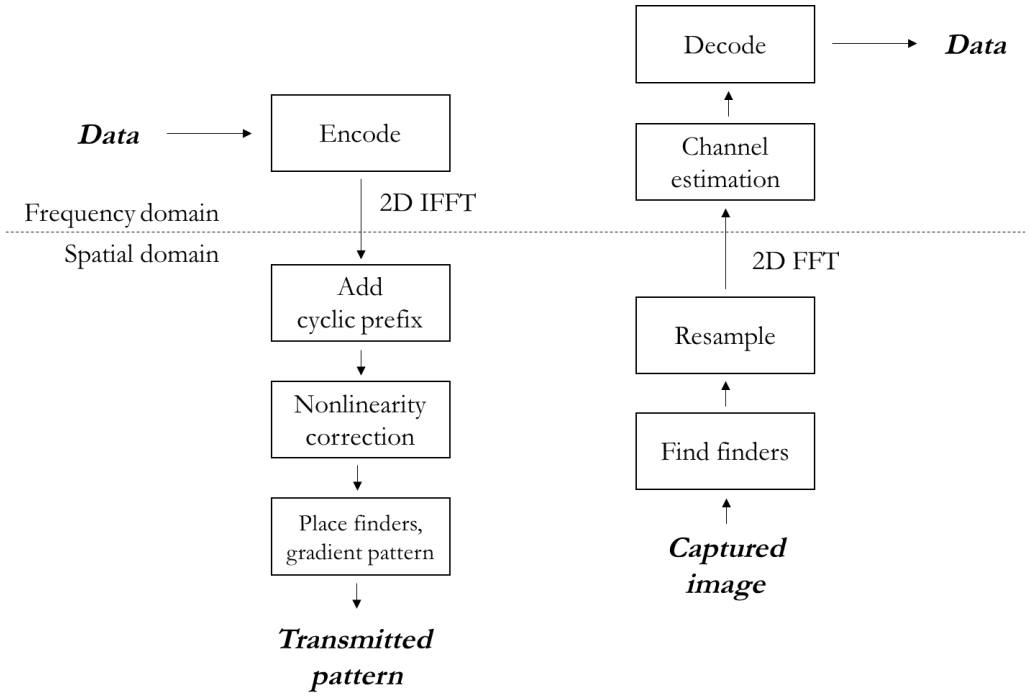



Figure 4.1: System block diagram.

In this work, we make an effort to address the three challenges which degrade the capability of screen-to-camera communication: perspective distortion, blurring effect, and the nonlinearity between the transmitted pixel intensities and the received image intensities. We apply several techniques to both the transmitter and receiver side and analyze their effectiveness on improving the communication range. Figure 4.1 shows our system



doi:10.6342/NTU201903432

block diagram.

On the transmitter side, first, we encode the data in frequency domain. Next, 2D IFFT is applied to transform the data from frequency domain to spatial domain to create a transmitted code. Then, to accurately locate the transmitted code in the observed scene, four finder patterns are added. In addition, in order to perform nonlinearity calibration, we add four gradient patterns to the transmitted code.

After the transmitted code is shown on the screen, the camera would capture the code and calculate nonlinearity relationship between the screen and the camera. Based on the calculated result, nonlinearity calibration algorithm can be performed at either the transmitter side or the receiver side.

On the receiver side, we divide the decoding process into four stages to deal with the captured image. First, we determine the position of the transmitted code by locating the location of the four finder patterns. Then, we estimate the frequency phase shift of the known data inserted in the transmitted code. With the estimated phase shift, we are able to restore the transmitted code from the perspective distortion by resampling the transmitted code and retrieve the intensity value from the correct locations. Eventually, the Fourier transform is applied to obtain the data in frequency domain.

## 4.2 Transmitter

### 4.2.1 Design of transmitted code

Figure 4.2 1 shows an illustration of our designed code. The transmitted code can be divided into three segments. The largest region located in the middle of the code is called a data region, where the transmitted data are encoded. A data region is composed of 4 subregions (shown in Figure 4.3). The reason for 4 subregions will be described in Chapter 4.3.2.

At the corners of the data region are the four finder patterns, which are used for identifying and locating the transmitted code in a captured image. To ensure the uniqueness of the finder patterns, the design of our finder patterns is according to the position pat-



doi:10.6342/NTU201903432

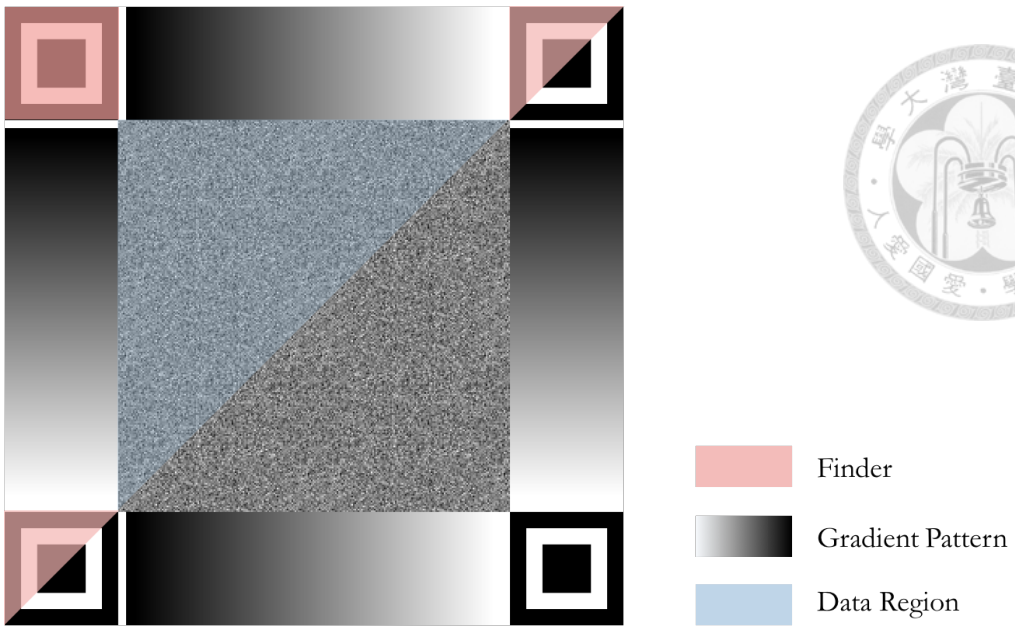



Figure 4.2: An example of transmitted code.

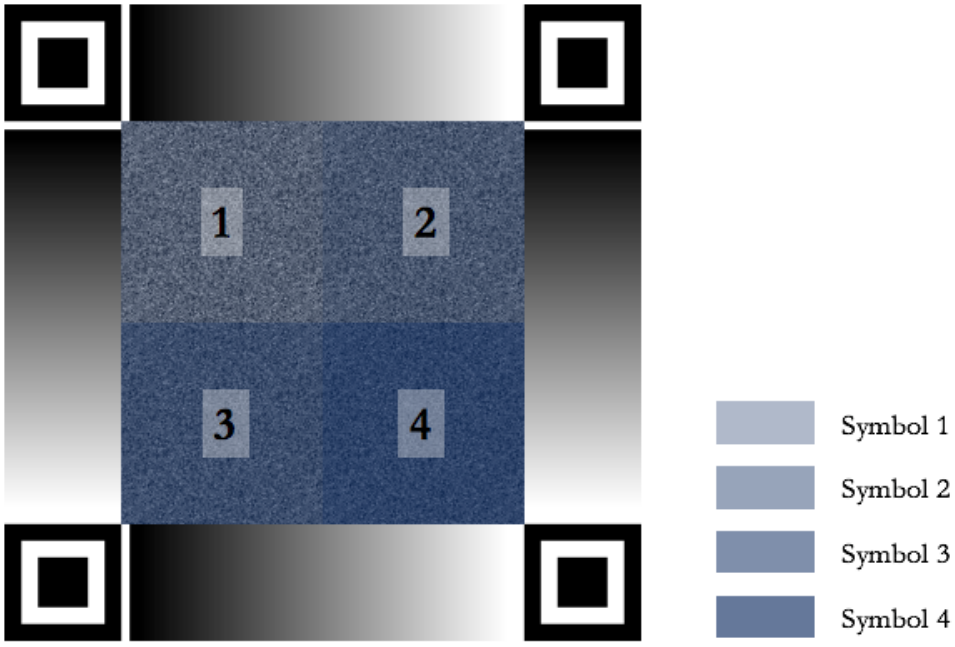



Figure 4.3: An illustration of the four symbols in one data region.

terns used in QR code where the ratio of the width of the color block is 1:1:3:1:1 [12], as illustrated in Figure 4.4.

To calibrate the nonlinear relationship of luminance between the transmitter and receiver, there are four gradient patterns in the transmitted code whose intensity values range from 0 to 255 calibrate the transmitted code. In addition, to prevent the dark blocks in the



doi:10.6342/NTU201903432

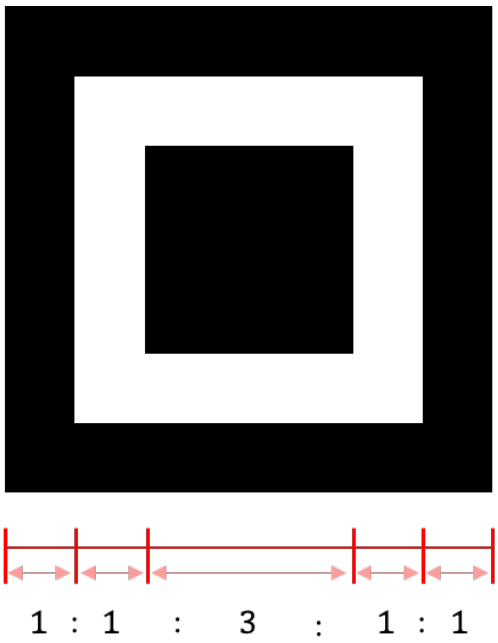



Figure 4.4: An illustration of the finder pattern.

finder pattern from being blended with the gradient pattern, extra spacing is applied.

### 4.2.2 Code generation

In our system, we consider a data subregion as an OFDM symbol, and thus a data region is composed of four OFDM symbols as a macro symbol. The main reason for this design comes from [11], where the four symbols are used to estimate the shift of the four corners caused by perspective distortion, respectively. The correction algorithm will be detailed in Chapter 4.3.2.

For each symbol, the transmitted data is first modulated with phase shift keying (PSK) and filled into a two dimensional matrix with the center matrix value (i.e., the DC value) set to be zero. We also insert known data values act as pilots in the matrix for performing resampling algorithm and channel estimation. This data matrix is then fed to 2D inverse Fast Fourier Transform (IFFT) to create a two dimensional output. However, unlike traditional OFDM, which transmits both real and imaginary components, images only contain real intensity values. Therefore, it is important to ensure that the outputs of 2D IFFT are real numbers to avoid data loss during the transformation. Fortunately, this issue can be solved by forming the transmitted data into a two dimensional Hermitian matrix. A Her-



doi:10.6342/NTU201903432

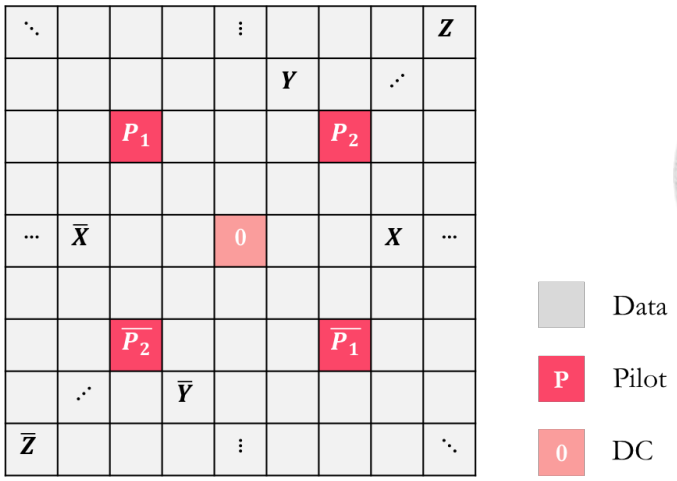



Figure 4.5: An OFDM symbol.

mitian matrix is a complex square matrix whose conjugate transpose is equal to itself. Thanks to the special property of Hermitian matrix, the output from 2D IFFT of a Hermitian matrix is entirely real. Figure 4.5 shows an example of one sub-symbol (i.e., one subregion) with transmitted data, pilot carriers and DC value set to be 0 in the center.

In addition, to prevent the symbols from being affected by the adjacent symbols or the background, we add a cyclic prefix (CP) to each symbols. As shown in Figure 4.6, we adopt the same scheme used in [11] which is to append the CP around the symbol by copying the first $n$ rows from the top of the symbol to the bottom of the symbol and vice versa. With the use of the CP, we can not only lower the interference from the adjacent symbols but also reduce the error caused by the imprecise sampling.

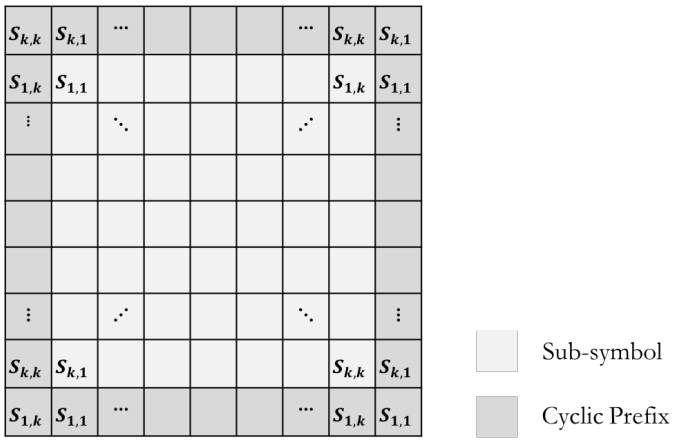



Figure 4.6: An OFDM symbol with CP.



doi:10.6342/NTU201903432

### 4.2.3 Nonlinearity calibration

In a screen-to-camera system, the channel between the transmitter (i.e., the screen) and the receiver (i.e., the camera) is considered to be nonlinear, which means the received image intensity is not linearly proportional to the transmitted intensity. The main reason for the nonlinearity is because that the received image intensity is determined by the transmitted intensity as well as the sensitivity or configurations of the camera. For instance, if the camera is configured to have a long exposure time, the received image intensity is more likely to be closer to 255. Whereas, if the camera has a short exposure time, the received intensity would be close to 0. In addition, the imperfectness of the display screen would further aggravate the nonlinear channel, as the displayed image intensity might also be not linearly proportional to the original image intensity, i.e., the transmitted code generated by our program.

To eliminate the nonlinearity between the transmitted and received intensity, in this work, we design a gradient pattern to measure the relationship between the transmitted and received intensity. Four gradient patterns are placed between the four finder patterns with the range of intensity of [0, 255]. After we extract the gradient pattern from the captured image, we can obtain the image intensity received by the camera in the current camera setting and environment. Eventually, by mapping the received intensity to the transmitted intensity, we can create a look-up table called *responding curve* for nonlinearity calibration.

To realize nonlinearity calibration, a transmitted code is first captured, and we introduce two different ways for performing nonlinearity calibration on transmitter or receiver side:

- Nonlinearity calibration on transmitter side The nonlinearity calibration performed on the transmitter side requires a reversed communication link from the camera to the screen. After the receiver obtains the responding curve, it sends the responding curve to the transmitter. Then, the transmitter can adjust the intensity values of the transmitted code according to the responding curve and generate a new version of the transmitted code.



doi:10.6342/NTU201903432

- Nonlinearity calibration on receiver side Since there might not always be a reversed communication link from the camera to the screen, the nonlinearity calibration can also be performed on by recovering the original intensity values at the receiver side with less effectiveness. However, the main drawback of this approach is that if the received image is already interfered by the blurring effect, the recovered intensity is more likely to still be wrong.

## 4.3 Receiver

### 4.3.1 Code detection

The first step of decoding is to determine the location of the transmitted code in a captured image. To locate the finder patterns, first, we binarize the captured image with a global image threshold. Next, image processing techniques such as opening technique of morphology are used to reduce the noise interference. Then, we search the positions of the finder pattern in which the ratio of the dark and light blocks is 1:1:3:1:1 in both horizontal and vertical direction.

After locating the center of the finder patterns, we are able to estimate the positions of the corners of the data region. With the coordinates of the corners of the data region, we can sample the entire data region using 2D interpolation and retrieve the intensity values for decoding.

### 4.3.2 Resampling algorithm

Similar to traditional OFDM algorithm where the signal must be precisely sampled to avoid decoding error, sampling is the most crucial part in our decoding process, which dominates the decoding performance in our research. However, due to the various distance and the non-perpendicular viewing angle between the screen and the camera, we need to tackle two challenges which cause inaccurate sampling during the communication.

The first reason for inaccurate sampling is perspective distortion. We regard the pixel values on the screen as the signal samples at the transmitter, and the pixel values in the



doi:10.6342/NTU201903432

captured images as the signal sampled by the receiver. When the screen and the camera are not perfect aligned, there is a viewing angle leading to perspective distortion. According to [11], perspective distortion can be considered a sampling problem where different sampling rates are applied at different parts of the transmitted code. The part close to the camera would occupy larger space (i.e., more pixels) in the image, which can be seen as a higher sampling rate. On the other hand, the part far from the camera is smaller, i.e., occupies fewer pixels in the captured image, which refers to a lower sampling rate. This inconsistent sampling rate in spatial domain would lead to the phase shift in frequency domain, causing decoding error.

The second reason is inaccurate positioning of the data region caused by blurring effect. The blurring effect introduces blurred edges to the transmitted code. Blurred edges further increase the difficulty in distinguishing between the finder patterns and the background, making it harder to locate the corners of the data region precisely. The offset between the calculated and actual location of the data region's corners would also cause phase shifts in frequency domain. To decrease the symbol error caused by the phase shift, it is important to resample the data region of the transmitted code as precisely as possible. To address this issue, we need to correct the location of the data region's corners.

In this research, we design a resampling method based on the solution of the perspective correction in Pixnet [11]. We treat our problem as the sampling problem which is similar to sampling frequency offset (SFO) in OFDM. In traditional OFDM channel, SFO occurs when the transmitter and receiver have inconsistent sampling rate. The difference of sampling rate between the transmitter and the receiver causes the phase shift in the frequency domain, which is proportional to the frequency. According to the time-shifting property of the Fourier transform, the phase shift in the frequency domain can be translated into the time shift in the time domain. PixNet generalized this time-shifting property to 2D Fourier transform, while the time shift is now the pixel offset in the spatial domain. Since there are two directions in the spatial domain (i.e., horizontal and vertical direction), the pixel offset of both directions need to be estimated, as shown in Figure 4.7. The pixel



doi:10.6342/NTU201903432

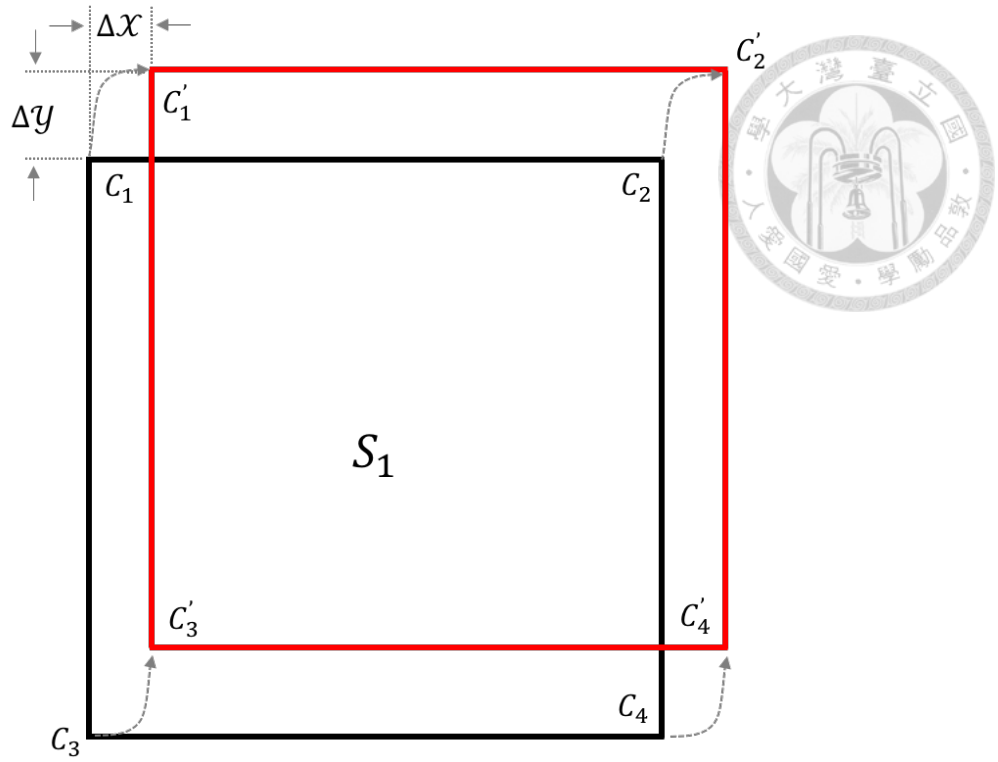



Figure 4.7: An OFDM symbol with constant shift.

offsets can be obtained from the phase shift $\Delta\theta_x$ and $\Delta\theta_y$:

$$(\Delta x, \Delta y) = \left( \frac{L_s \Delta\theta_x}{2\pi}, \frac{L_s \Delta\theta_y}{2\pi} \right), \tag{4.1}$$

where $L_s \times L_s$ is the size of the 2D FFT (i.e., the size of a symbol).

From this equation, we can see that the accuracy of the phase shift estimation plays an important role in determining the efficiency of the resampling algorithm. To this end, we design a novel method to estimate the phase shift. We first calculate the phase shifts experienced by each inserted pilots. Since the phase shift is proportional to the frequency, ideally, the phase shift of each pixel block can be fitted into a linear equation. Therefore, we can use a linear regression model to fit the phase shift of pilots. However, noise and interference in the real-world channel creates outliers which interferes the regression result. To avoid this, we instead apply the random sample consensus (RANSAC) algorithm to estimate the slope of phase shift. In RANSAC algorithm, the least squares method is adopted to calculate the distances between each point and the result model, which evaluates the performance of the selected models. Since there are x axis and y axis, we will



doi:10.6342/NTU201903432

estimate the slope of the phase shift of each axis, respectively. After obtain the slopes of the phase shift, we can apply these two values to equation 4.1 to obtain the pixel offset in each axis.

In conclusion, the process of the resampling algorithm can be divided into the following steps:

1. Use the pilots to calculate the phase shift in different frequencies.

2. Sort the set of the phase shifts according to the frequency in the x axis and y axis.

3. For each axis, estimate the slope of the phase shift by the RANSAC algorithm.

4. Translate the phase shift in the frequency to the pixel offset in the spatial domain.

According to [11], an OFDM symbol can be used to obtain only one pair of pixel offsets. However, as shown in Figure 4.8, the distortion can happen at the four corners. Thus, we need to estimate 4 pairs of pixel offsets. Therefore, we combine four symbols in one data region which is arranged into a two by two matrix. Eventually, we could estimate the pixel offset for four corners separately with the four symbols in data region using the following formulas:

$$\begin{pmatrix} \Delta\alpha_x \\ \Delta\beta_x \\ \Delta\gamma_x \\ \Delta\delta_x \end{pmatrix} = \frac{P_R}{L_R}\frac{16L_S}{2\pi} \begin{pmatrix} 9 & 3 & 3 & 1 \\ 3 & 9 & 1 & 3 \\ 3 & 1 & 9 & 3 \\ 1 & 3 & 3 & 9 \end{pmatrix}^{-1} \begin{pmatrix} \Delta\theta_{S_1,x} \\ \Delta\theta_{S_2,x} \\ \Delta\theta_{S_3,x} \\ \Delta\theta_{S_4,x} \end{pmatrix} \tag{4.2}$$

$$\begin{pmatrix} \Delta\alpha_y \\ \Delta\beta_y \\ \Delta\gamma_y \\ \Delta\delta_y \end{pmatrix} = \frac{P_R}{L_R}\frac{16L_S}{2\pi} \begin{pmatrix} 9 & 3 & 3 & 1 \\ 3 & 9 & 1 & 3 \\ 3 & 1 & 9 & 3 \\ 1 & 3 & 3 & 9 \end{pmatrix}^{-1} \begin{pmatrix} \Delta\theta_{S_1,y} \\ \Delta\theta_{S_2,y} \\ \Delta\theta_{S_3,y} \\ \Delta\theta_{S_4,y} \end{pmatrix} \tag{4.3}$$



doi:10.6342/NTU201903432

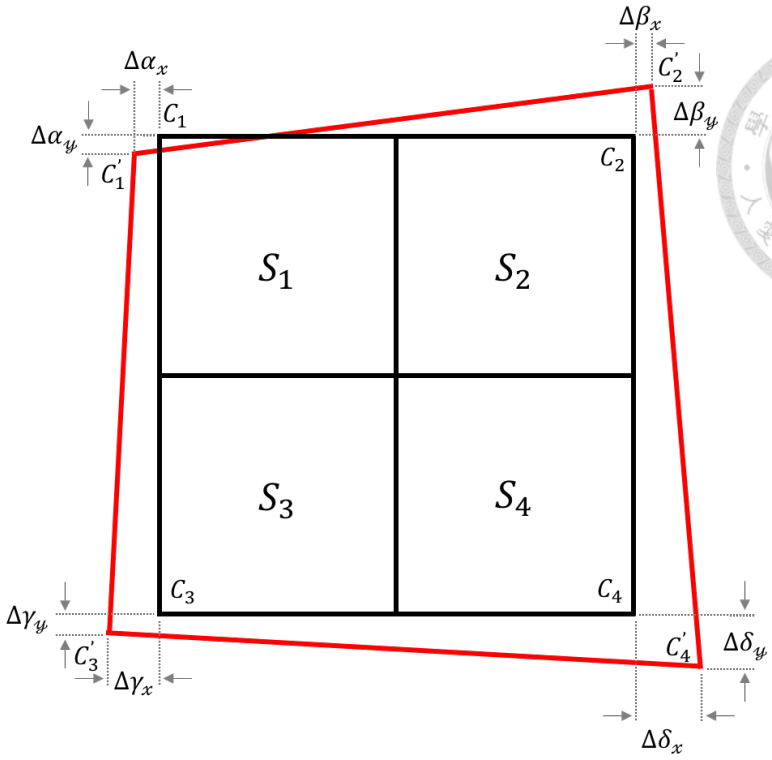



Figure 4.8: An OFDM symbol with distortion.

### 4.3.3 Remove channel effect (phase)

After the data region is accurately sampled, the intensity values in the data region are still highly polluted by the blurring effect. The blurring effect can be interpreted as a low pass filter which eliminates the significant details of the transmitted pattern, i.e., a loss of the high frequency content. As a result, the data in the high-frequency channels would be severely attenuated, leading to the decoding error in screen-to-camera communications.

In general, the blurring effect is mainly caused by the non-alignment between the pixel blocks in the transmitted code and the pixels in the image sensor. The non-alignment happens when the pixel cell of the image sensor observes the intensity of multiple pixel blocks in the transmitted code. As a pixel cell outputs the total intensity value within its sensing area, the output intensity would be the combination of all the pixel blocks observed. Non-alignment can happen at any distance. For example, at short distance, the non-alignment takes place when the image sensor observes the boundary of two transmitted pixels, which not only reduces the sharpness but also high frequency information of the transmitted image. On the other hand, when the distance between the transmitter and the camera is long,



doi:10.6342/NTU201903432

one pixel cell in the image sensor observes multiple transmitter pixels, outputting a summation of the observing pixel values.

In addition to the non-alignment, at long communication distance, the insufficient sharpness of the lens and the defocusing of the camera aggravate the influence of the blurring effect. This is due to that the imperfectness of the lens structure would make the incident light slightly scatter when passing through the lens, resulting in color mixing. Moreover, even the small movement of human body such as shaking or jittering when taking a picture would result in blurring as well.

To mitigate the blurring effect in the data region, in this work, we treat the blurring effect as another kind of sampling problem. But unlike the resampling method in which the phase shift in the frequency domain is translated into the pixel offset in the spatial domain, we estimate the channel caused by the blurring effect in the frequency domain. To do so, we take advantage of the pilots inserted in the transmitted code as mentioned before. Then, we can obtain the phase shift difference between the transmitted data and received data. We calculate blurring effect channel by fitting it to a two dimensional plane. Finally, the received data could be restored by dividing the estimated channel expressed as the following equation:



doi:10.6342/NTU201903432

# Chapter 5

# Implementation

## 5.1 Encoder

We use a Matlab program to generate our transmitted code. The algorithm can be divided in four steps:

1. Modulate the transmitted data with PSK.

2. Insert the modulated data and pilots into the upper half of a square matrix, and insert the complex conjugate form of the data in the lower half of the matrix to form a Hermitian matrix.

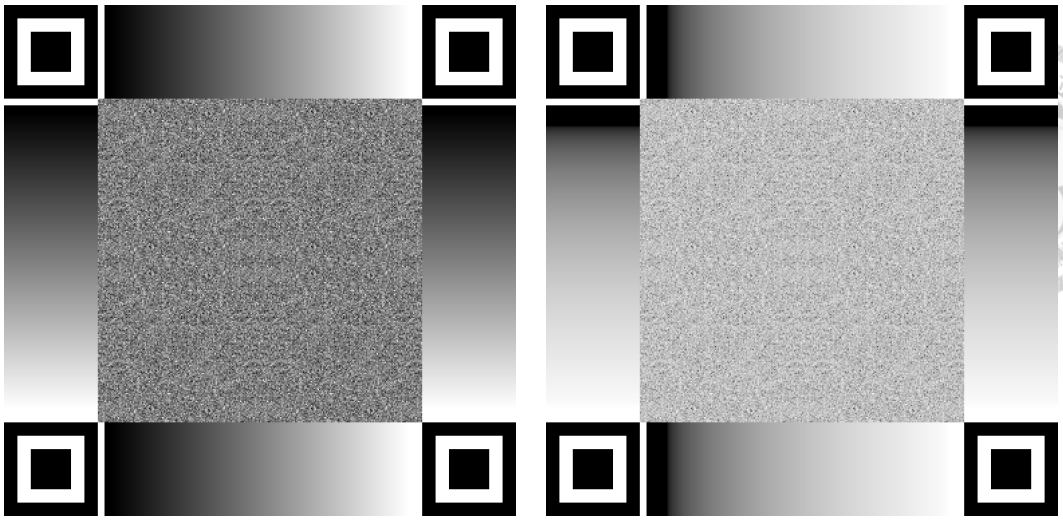
3. Apply 2D IFFT to the data matrix to obtain the intensity values in spatial domain. Since the result from 2D IFFT ranges from -1 to 1, we need to map the result to 0 to 1 for showing the image.

4. Finder and gradient patterns are added to create a complete transmitted code, as shown in Figure 5.1(a).

To perform nonlinearity calibration, the transmitter obtain the intensity curve calculated by the receiver. According to the intensity curve, the transmitter adjusts the pixel intensity of the transmitted code. As an example, Figure 5.1(b) shows an example of the transmitted code after nonlinearity calibration.



doi:10.6342/NTU201903432



(a) Before nonlinearity calibration. (b) After nonlinearity calibration.

Figure 5.1: An example of nonlinearity calibration.

## 5.2 Decoder

After we captured the transmitted code with an image, we need to find the finder patterns to locate the data region in the captured image. To enhance the efficiency and accuracy, we apply some image processing technique to first filter out the noise. The steps are detailed as follows:

1. Due to the format of the captured image, which is BMP for the Flea3, we need to translate the image from RGB color space to grayscale space. Then, we binarize the grayscale image with the use of Otsu's method to obtain a proper threshold.

2. The Bayer pattern in the image sensor would cause uneven exposure in the captured image, resulting in a jagged edge of the contour of the finder patterns. To remove the noise, we apply the opening technique in morphological image processing to restore the original contour of the finder pattern.

3. To compute the location of the four finder patterns in the binary image, we search the positions of the finder pattern in which the ratio of the dark and light blocks is 1:1:3:1:1 in both horizontal and vertical direction. Then, the center of the finder patterns could be located.

4. According to the center of four finder patterns, we could locate the position of the data region and calculate the sampling point by 2D interpolation.



doi:10.6342/NTU201903432

After finishing the above steps, the sampling data would be transformed into the frequency domain by 2D FFT. Resampling method and channel removal will be applied to adjust the sampling point and restore the erroneous bits. Eventually, the data would be decoded from the transmitted code.



doi:10.6342/NTU201903432

# Chapter 6

# Evaluation

In this chapter, firstly, we describe the experimental setup. Then, we introduce the key ratio which is used to enable the fair comparison among multiple transmitters for evaluating the system performance. Finally, we analyze the effectiveness of our proposed methods based on the bit error rate (BER) at different communication distances.

## 6.1 Evaluation environment

**Hardware:** For the transmitter, three different screens are used to display the transmitted code: a 24" flat display Dell P2416d、iPhone 6+ which has a 5.5" display and Apple watch series 4 with 1.8" display. The comparison of these three transmitters is shown in Table 6.1. For the receiver, to understand how the resolution of a camera and the quality of the image sensor affect the performance of the our proposed solutions, two cameras with different resolutions and image sensors are used in the experiments: a Point-Grey Flea3 (RL3-U3-88S2C-C) and a Canon 70D is used. The comparison of these two cameras is listed in Table 6.2. We first select a lens with 16 mm focal length for Flea 3. In order to maintain a similar field of view (FOV) between these two different type of cameras, we choose the lens whose the focal length is determined according to the following equation:

$$\alpha = 2arctan\frac{d}{2f}, \tag{6.1}$$



doi:10.6342/NTU201903432

where $\alpha$ represents the angle of FOV, $d$ represents the size of the image sensor and $f$ is the focal length of the adopted lens. By substituting the image sensor size of Canon 70D into this equation, we can derive the focal length for Canon 70D, which is 50 mm.

Table 6.1: Transmitter specifications

| | DELL P2416d | iPhone 6+ | Apple watch series 4 |
|---|---|---|---|
| Resolution | 2560*1440 | 2560*1440 | 448*368 |
| Pixel Pitch ($mm$) | 0.205*0.205 | 0.063*0.063 | Not found |
| Brightness ($cd/m^2$) | 300 | 566 | Not found |
| Contrast Ratio | 1000:1 | 1300:1 | Not found |

Table 6.2: Receiver specifications

| | Flea3 | Canon 70D |
|---|---|---|
| Image Sensor Format | Sony IMX121 (1 / 2.5") | Canon APS-C |
| Type of Sensor | CMOS | CMOS |
| Resolution | 4096*2160 | 5472*3648 |
| Pixel Size ($\mu m$) | 1.55*1.55 | 4.1*4.1 |

**Symbol size of the transmitted code:** For Dell P2416D and iPhone 6 plus, the symbol size is set to be 101, corresponding to a data region of 242x242 pixels. For Apple watch, since its display is much smaller than the former ones, we use a symbol size of 41, referring to a 102x102 pixels data region. On the other hand, since the luminance level of the transmitter greatly influences the system performance, we set the luminance level of the transmitter to the maximum level. In addition, as the pixel size of the transmitter is too small, resulting in low luminance output, we combine 2x2 pixels as a macro pixel for displaying the transmitted code.

**Performance metric:** Most of the past works evaluate their performance with the use of BER at different communication distances. The main problem of this evaluation method, however, is that we cannot compare the performance of the same transmitted code on different transmitters, as the transmitted code on a screen with small pixel pitch is indeed smaller than that displayed on a screen with bigger pixel pitch. Since the size of displayed transmitted code has a significant influence on the size of the received image, at a same communication distance, the decoding performance with a smaller transmitted code would be worse than a bigger one. To enable a fair comparison across various transmitters,



doi:10.6342/NTU201903432

in this thesis, we define a ratio of the size of received data region and the size of the transmitted data region as *magnification ratio* which can be expressed as:

$$magnificationratio = \frac{P_r}{P_t}, \tag{6.2}$$

where $P_r$ is the size of the received data region in pixel and $P_t$ is the size of the transmitted data region in pixel. As the observed size of the displayed transmitted code decreases when the distance between the transmitter and the receiver increases, for the same transmitter, the smaller magnification ratio represents that distance between the transmitter and the receiver is longer.

**Experimental setup:** In the experiments, we setup a screen-to-camera link in an indoor and static scenario, as shown in Figure 6.2. Since the size of the transmitters are different, to maintain a similar magnification for comparison, the system performance of each transmitters are measured at different scales of distance.

1. Dell P2416D (Figure 6.1(b)): We measure the system performance using Dell monitor with a magnification ratio ranging from about 4 to 0.8, corresponding to a distance of 1.15$\sim$4.4 m. The error rate is measured for every 20 cm.

2. iPhone 6+ (Figure 6.1(c)): We measure the system performance using iPhone with a magnification ratio ranging from about 3.8 to 1.1, corresponding to a distance of 0.43$\sim$1.8 m. The error rate is measured for every 10 cm.

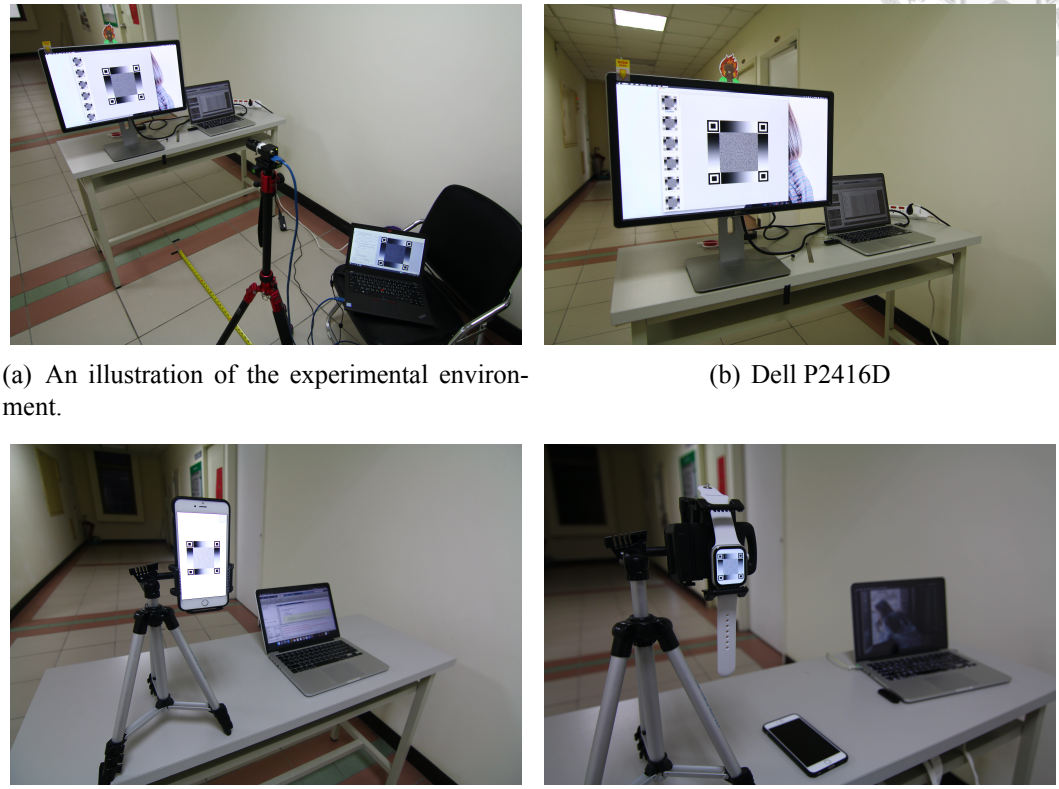
3. Apple watch (Figure 6.1(d)): We measure the system performance using Apple watch with a magnification ratio ranging from about 4 to 0.8, corresponding to a distance of 0.54$\sim$1.6 m. The error rate is measured for every 10 cm.

At each distance, the transmitter is placed in the center of the FOV of the receiving camera. To enable the maximum system capability, we develop an algorithm to ensure the exposure setting of the camera is configured to have the best receiving performance (i.e., the maximum dynamic range) under the observed scene for each transmitter and receiver pair. To do so, we select several exposure settings and measure their responding curve of the intensity values between the screen and the camera with the obtained gradient patterns. The dynamic range is defined as the difference between the maximum and the minimum



doi:10.6342/NTU201903432

value of the responding curve. Among the measurement result, the setting with the largest dynamic range would be selected as our camera exposure setting. Three kinds of symbols modulated with BPSK, QPSK, and 8PSK are tested in our experiments.



(a) An illustration of the experimental environment.

(b) Dell P2416D

(c) iPhone 6+

(d) Apple watch series 4

Figure 6.1: Experimental environment.

## 6.2 System performance when using different transmitters

In the following, we first report the system performance of different transmitters when using different modulation schemes. Next, we show the results with and without nonlinearity calibration, resampling and channel estimation. Then, we compare the system performance under different receivers. Finally, the system performance under different viewing angles is presented.



doi:10.6342/NTU201903432

### 6.2.1 Modulation

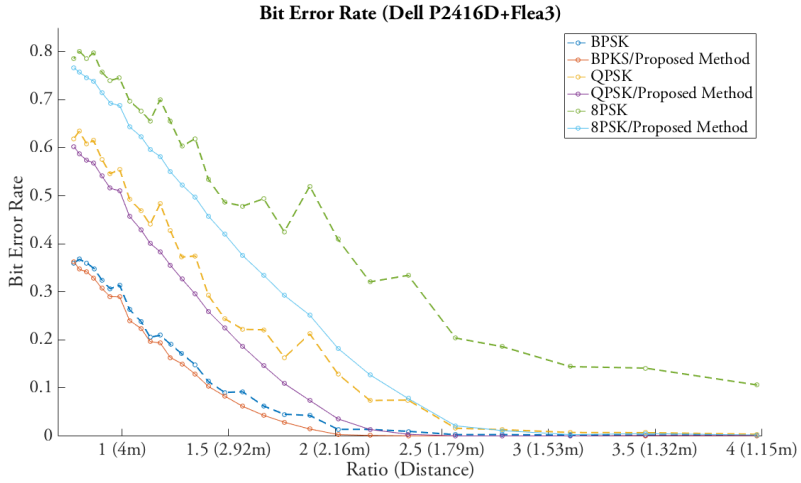
We first report the overall system performance of the three modulation schemes with the use of three different transmitters. As shown in Figure 6.2, for all of the three transmitters, our proposed method is able to reduces the error rate. Especially, significant improvement has been made to high-order modulations (8PSK in our case). Nevertheless, since 8PSK has the smallest error margin among the three modulation schemes and thus the lowest fault tolerance, its error rate is slightly higher at all ratio. On the other hand, BPSK, as one of the simplest modulation scheme, can maintain a low error rate even at a ratio of 1.5.

In addition, as one can see, the error rate increases dramatically when the ratio is less than 1.5. This is because of the blurring effect is more severe at longer communication distance. As mentioned before, the blurring effect is similar to the low-pass filter which has the 3x3 or 5x5 kernel. Therefore, when the transmitting distance increases, data in high frequency is interfered by the blurring effect causing serious attenuation. Moreover, the longer communication distance suffers from the Inverse-square law, where each pixel cell in the image sensor could only receive a smaller intensity value, resulting in the reduced dynamic range of the observed transmitted code. The reduced dynamic range would lower the range of the intensity to display the brightness.

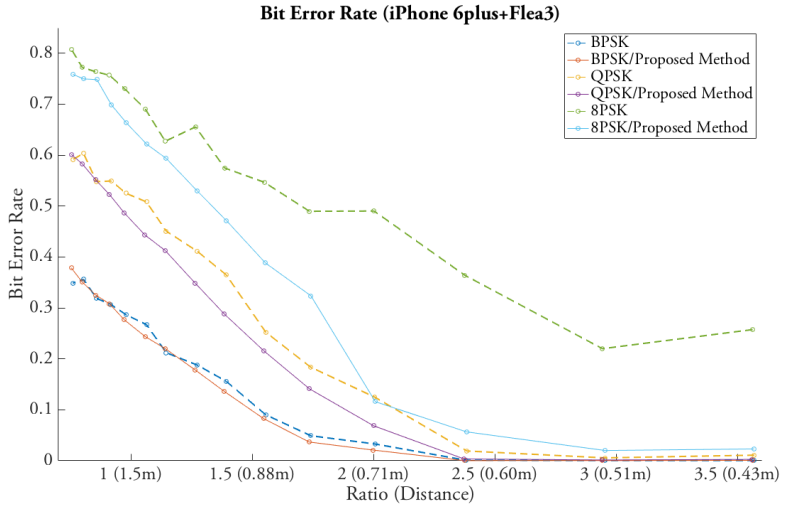
In addition to the BER versus magnification ratio, we further analyze the error distribution of each modulation scheme. Figure 6.3 are the error distribution of BPSK, QPSK, and 8PSK obtained at different ratios. We can see that most of the errors take place at the outer part of the image, which corresponding to the high frequency band introduced in the previous chapters. As the distance increases, the errors spread to medium frequency. This is because the blurring effect becomes stronger at longer distance, therefore, introducing a stronger low pass filter which attenuates wider frequency bands. Fortunately, the data in low frequency remains even when the ratio is down to almost 1.



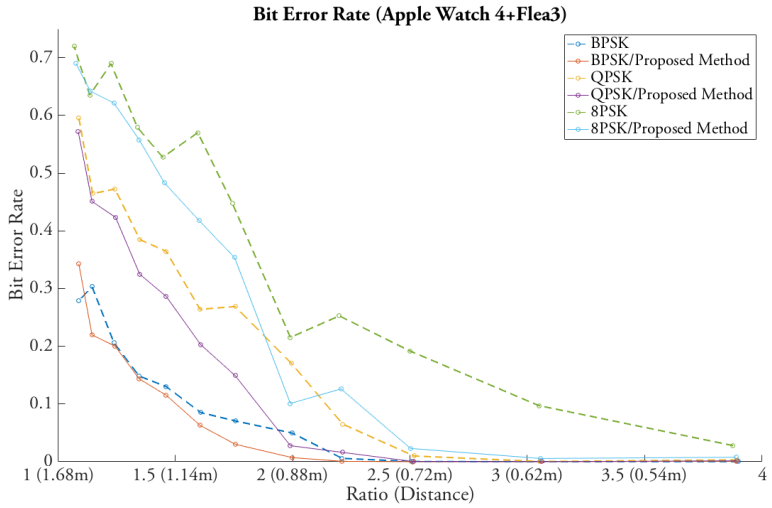
doi:10.6342/NTU201903432




(a) Dell P2416D




(b) iPhone 6+

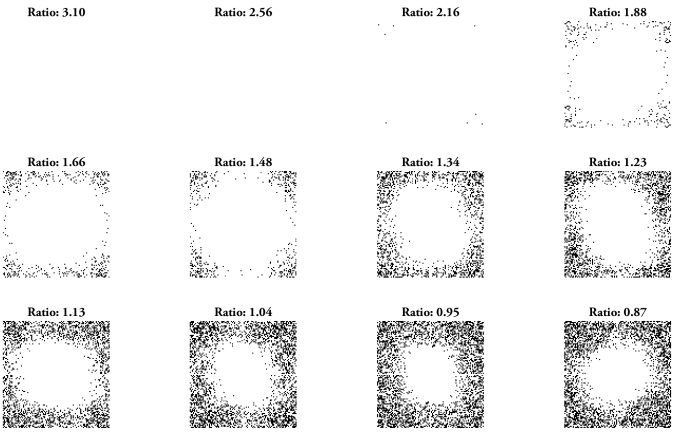



(c) Apple watch series 4

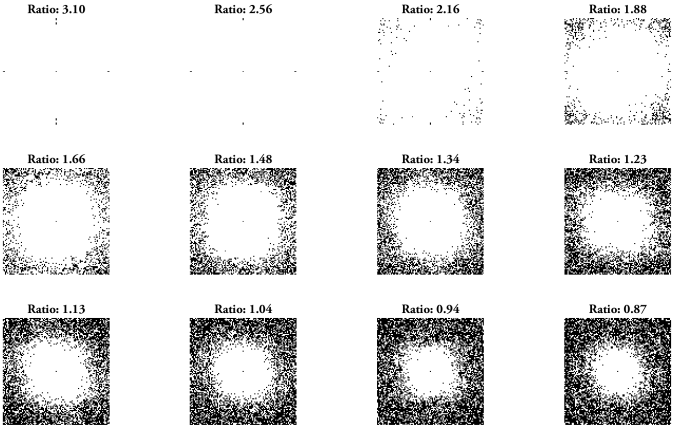
Figure 6.2: Comparison of BER with and without the proposed methods.



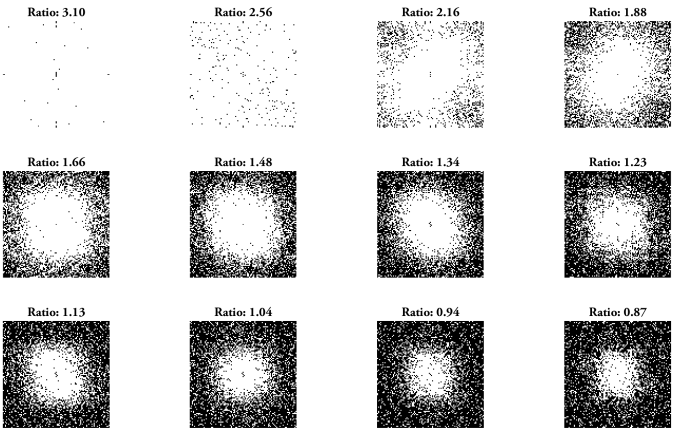
doi:10.6342/NTU201903432




(a) Dell P2416D




(b) iPhone 6+




(c) Apple watch series 4

Figure 6.3: Error distribution of BPSK, QPSK and 8PSK.



doi:10.6342/NTU201903432

### 6.2.2 Effectiveness of each proposed methods

Here we compare the effect of each of our proposed methods. Figure 6.4 is the BER of 8PSK at different ratios. We can see that even though the nonlinearity calibration greatly reduces the BER, but it cannot reduce the unsteadiness of the BER performance. This is because that although we try our best to put the transmitted code in the middle of the FOV of the camera, the viewing angle may not be perfect perpendicular, causing perspective distortion to the obtained transmitted code.

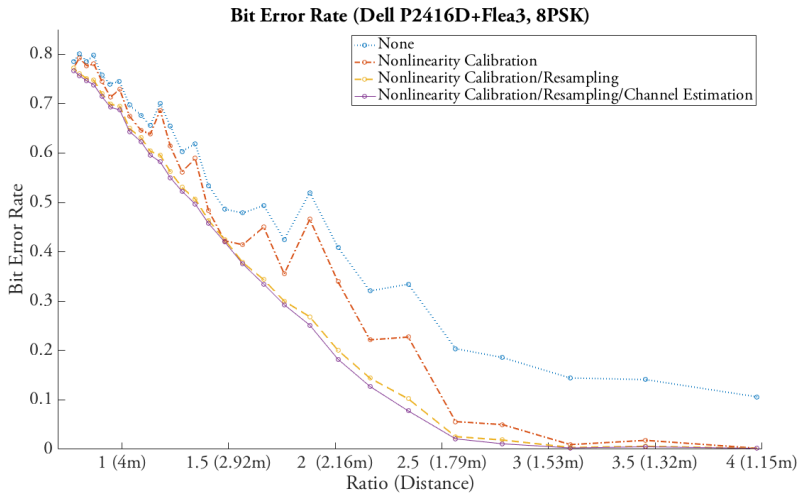
As one can see, this unsteadiness can be eliminated by the resampling method. In the three different scales of screens, resampling method decreases the inaccurate sampling which smooth the curve of BER. However, channel removal has limited effect on improving the BER when the resampling has been applied. This might be due to that the resampling method and channel removal both use the pilot to correct the phase shift, which introduce the similar effect on improving the communication performance. Moreover, we notice that the error rate is similar at the same ratio even with the different screen-camera pairs. The results show that the blurring effect has the less relevance between the type of screen.

### 6.2.3 Comparison between the resampling algorithm and channel removal

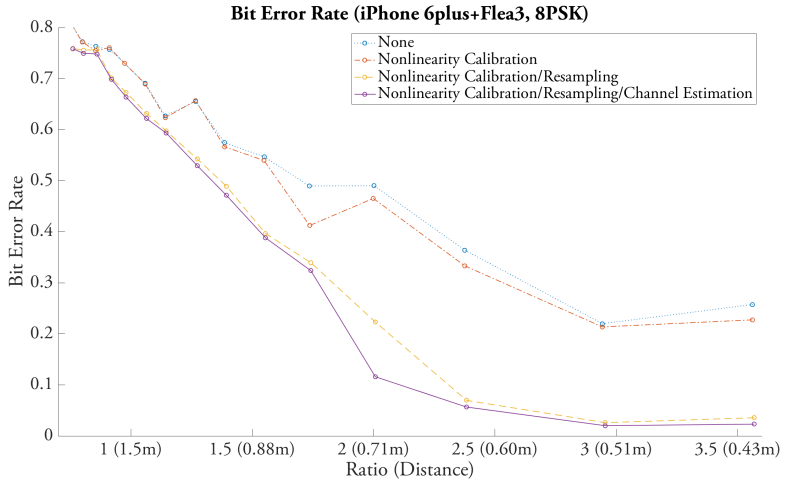
The results of previous section indicate that resampling algorithm and channel removal have the similar effect on improving the communication performance. To investigate this phenomenon, in this section, we compare the result of resampling and channel removal with different order of modulations and analyze the difference between these two methods. Figure 6.5 shows that generally, the resampling algorithm has the better effectiveness. The reason is that resampling algorithm calculates the accurate location of sampling point, which contains the more information data in high frequency. Whereas, channel removal could only correct the data using the phase shift of frequency which is obtained from the calculated channel.



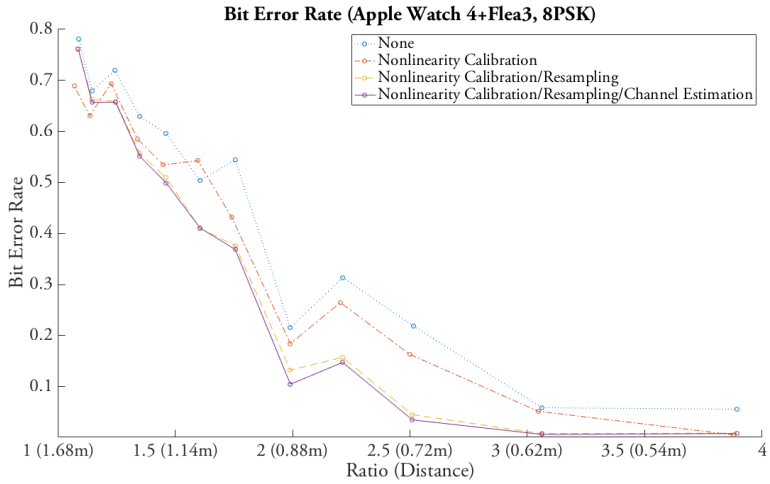
doi:10.6342/NTU201903432




(a) Dell P2416D




(b) iPhone 6+

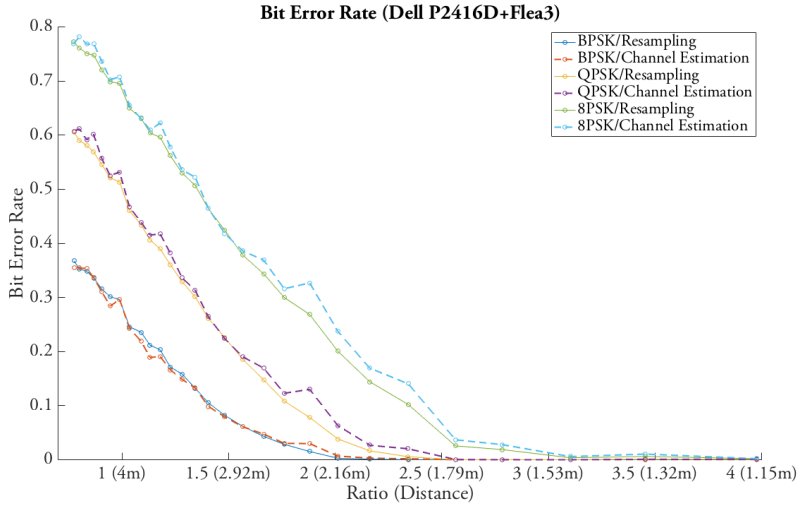



(c) Apple watch series 4

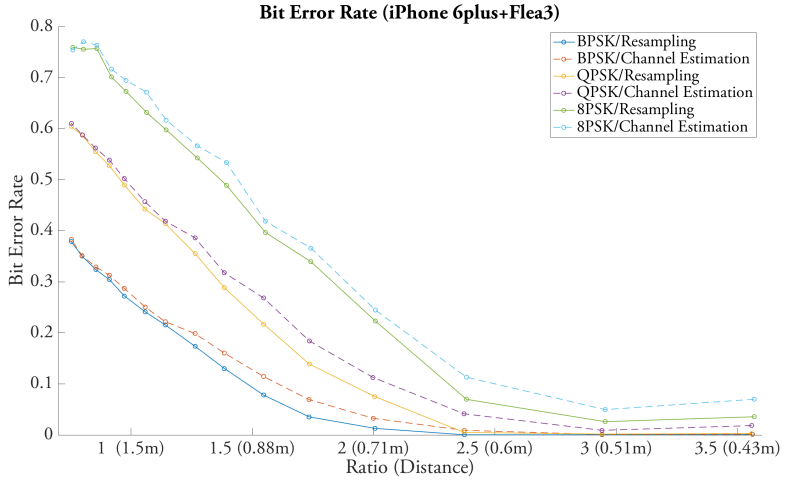
Figure 6.4: Error distribution of BPSK, QPSK and 8PSK.



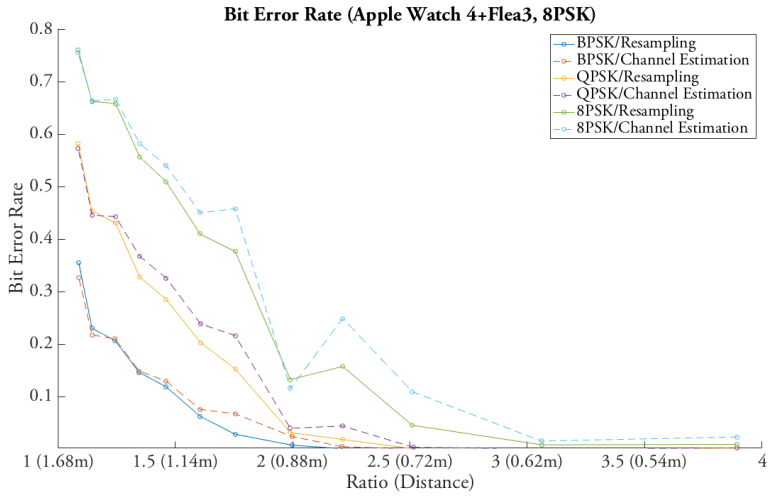
doi:10.6342/NTU201903432




(a) Dell P2416D




(b) iPhone 6+




(c) Apple watch series 4

Figure 6.5: Comparison between resampling algorithm and channel removal.



doi:10.6342/NTU201903432

## 6.3 System performance when using different receivers

These section shows the result of using different cameras. Theoretically, the ability of image sensor determines the quality of the communication channel. Figure 6.6 shows that nonlinearity calibration has the better performance on Flea3. The reason is that the responding curve of the Canon 70D is much linear that Flea3 (as illustrated in Figure 6.7), and hence the nonlinearity problem is not that serious in Canon 70D than that in Flea3.

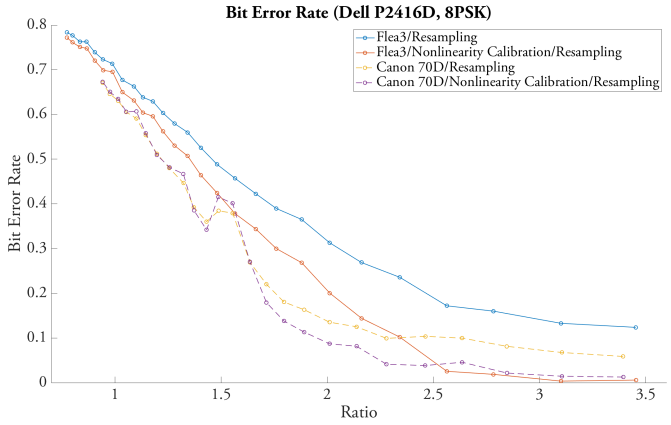



Figure 6.6: BER performance with different cameras.

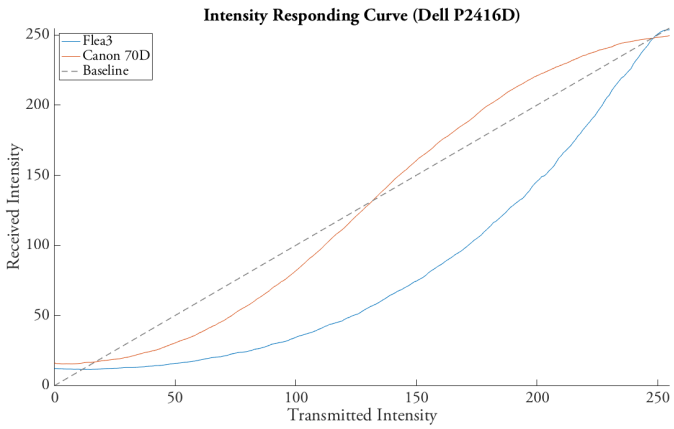



Figure 6.7: Responding curves of Flea3 and Canon 70D.



doi:10.6342/NTU201903432

# Chapter 7

# Conclusion and Future Work

Screen-to-camera communications have become one of the most popular ways for real-time data transmission in recent years, due to its low cost and convenience There are two kinds of encoding schemes to transform the transmitted data into images, spatial encoding and frequency encoding. For example, QR code is the most common communication scheme using spatial encoding. However, spatial encoding scheme tends to have short communication range and is vulnerable to blurring effect, which limits the capacity of screen-to-camera communications for handling various applications . On the other hand, frequency encoding features better resilience to blurring effect, and is thus chosen as the main encoding method to implement screen-to-camera communication in this work.

In this thesis, we aim to increase the working distance of screen-to-camera communication. We first investigate the main reasons that degrade the performance of screen-to-camera communications at long range and then propose several methods to address the degrading causes. Firstly, we design a nonlinearity calibration algorithm to minimize the difference between the intensity displayed on the screen and that received by a camera. Nonlinearity calibration allows the camera to capture a more accurate intensity value which is closer to the original transmitted intensity of the pixel block. Secondly, we improve the resampling method for correcting the perspective distortion. We apply the RANSAC algorithm to estimate the phase shift in the spatial frequency. The reason for using RANSAC is that it is more resilient to the outliers during the estimation and thus is able to further improve the accuracy of the resampling method. Finally, channel



doi:10.6342/NTU201903432

estimation has been applied to calculate the blur model.

We carry out experiments on our proposed method using three different types of transmitters and two kinds of receivers. The experimental results show that nonlinearity calibration algorithm significantly reduces the error rate by at most 20% in high-order modulation schemes. In addition, the error rate without using the resampling method is more varied due to incorrect sampling. This indicates that the resampling method is necessary for enabling accurate and robust sampling. However, in our experimental results, channel removal only has limited effect after applying the resampling. This may due to that most of the phase shift is corrected by the resampling algorithm, so there is not much phase shift left for channel removal.



doi:10.6342/NTU201903432

# Bibliography


[1] Iso. information technology—automatic identification and data capture techniques—data matrix bar code symbology specification. number iso/iec 16022:2006. geneva, switzerland, 2006.

[2] Iso. information technology –automatic identification and data capture techniques – aztec code bar code symbology specification. number iso/iec 24778:2008. geneva, switzerland, 2008.

[3] E. C. for Standardization. Automatic identification and data capture techniques –qr code 2005 bar code symbology specification: Iso/iec 18004-2006, 2016.

[4] T. Hao, R. Zhou, and G. Xing. Cobra: Color barcode streaming for smartphone systems. In *Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services*, MobiSys '12, pages 85–98, New York, NY, USA, 2012. ACM.

[5] F. Hermans, L. McNamara, G. Sörös, C. Rohner, T. Voigt, and E. Ngai. Focus: Robust visual codes for everyone. In *Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys '16, pages 319–332, New York, NY, USA, 2016. ACM.

[6] S. Hranilovic and F. R. Kschischang. A pixelated mimo wireless optical communication system. *IEEE Journal of Selected Topics in Quantum Electronics*, 12(4):859–874, July 2006.




doi:10.6342/NTU201903432

[7] W. Hu, J. Mao, Z. Huang, Y. Xue, J. She, K. Bian, and G. Shen. Strata: Layered coding for scalable visual communication. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 79–90, New York, NY, USA, 2014. ACM.

[8] M. Inc. The next-generation warehouse —long range scanning and emergence of 2d bar codes. technical report, 2011.

[9] T.-W. Kan, C.-H. Teng, and W.-S. Chou. Applying qr code in augmented reality applications. In *Proceedings of the 8th International Conference on Virtual Reality Continuum and Its Applications in Industry*, VRCAI '09, pages 253–257, New York, NY, USA, 2009. ACM.

[10] D. Parikh and G. Jancke. Localization and segmentation of a 2d high capacity color barcode. In *2008 IEEE Workshop on Applications of Computer Vision*, pages 1–6, Jan 2008.

[11] S. D. Perli, N. Ahmed, and D. Katabi. Pixnet: Interference-free wireless links using lcd-camera pairs. In *Proceedings of the Sixteenth Annual International Conference on Mobile Computing and Networking*, MobiCom '10, pages 137–148, New York, NY, USA, 2010. ACM.

[12] S. Tiwari. An introduction to qr code technology. In *2016 International Conference on Information Technology (ICIT)*, pages 39–44, Dec 2016.

[13] Z. Yang, H. Xu, J. Deng, C. C. Loy, and W. C. Lau. Robust and fast decoding of high-capacity color qr codes for mobile applications. volume 27, pages 6093–6108, Dec 2018.



doi:10.6342/NTU201903432