Title
Indoor Positioning System Using Visible Light Communication and Smartphone With Rolling Shutter Camera

Permalink
https://escholarship.org/uc/item/4280x06f

Author
Li, Zening

Publication Date
2016

Peer reviewed|Thesis/dissertation
Indoor Positioning System Using Visible Light Communication and Smartphone with Rolling Shutter Camera

A Dissertation submitted in partial satisfaction of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Zening Li

December 2016

Dissertation Committee:

Professor Gang Chen, Co-Chairperson
Professor Albert Wang, Co-Chairperson
Professor Ming Liu
The Dissertation of Zening Li is approved:

Co-Chairperson

Co-Chairperson

University of California, Riverside
Acknowledgments

First and foremost, I would like to express my special appreciation and sincere gratitude to my advisers, Prof. Gang Chen and Prof. Albert Wang, you are tremendous mentor to me in both my research as well as my life. Thanks for your continuous support, patient guidance and encouragement during the past four years. Besides, I want to thank Professor Ming Liu, for being my committee member and his insightful comments and feedback to my research.

I would express my special thanks to HongSheng Yu and Xing Zheng, for their great help to me about the Kalman filter part. Without them, I could not finish this work. Thanks Panruo Wu, Dongfang Zheng and Mingyang Li, who spent hours to discuss my research with me.

I want to thank Prof. Gang Chen and Adrienne Thomas for providing me this opportunity to study at UC Riverside. I thank my fellow labmates, LinChao Liao and Tian Lang, for research discussing and group experiments, for the sleepless nights we were working together before the deadlines, and for all the fun we have had in the last four years. I will never forget the time we spent together on desert in Palm Spring for outdoor experiment.

I am deeply grateful for Dr. Gang Chen and his wife Feng bo, Tian Lang and Pengchen Zhao for their care and encouragement after I had my ACL reconstruction surgery. And I am so fortunate to have my friends Linchao Liao, Panruo Wu, Peng Deng and Tian Lang at UCR to share all my good times and bad times with them. I enjoy the old times playing basketball with my friend Chi Tang, that’s how I tore my ACL. Thanks all of those
who helped me during my PhD. life.

I would like to thank my parents, for support me spiritually and physically for my whole life. Last but not the least, special thanks to my girlfriend Qi Wu, for all the happiness you brought to me during my PhD. life and thanks for your support and understanding of everything.
To my family and Yoyo for all the support.
Indoor positioning systems provide location based service within buildings. Because the Global Positioning System is usually unavailable in indoor environment, other positioning technologies making use of optical, radio or even acoustic techniques are used for indoor positioning application. In this dissertation, an indoor visible light communication positioning system using a smartphone with rolling shutter camera is proposed. The LED transmits periodical signals with different frequencies high enough to avoid flickering as its optical tags. The camera exploits the rolling shutter effect to detect the fundamental frequency of optical signals. This kind of systems use smartphone camera as the receiver with no requirement of extra hardware. And at the same time, the optical communication link allows a date rate much higher than the frame rates of traditional optical camera communication. For the work of this research topic carried out so far, the roles of camera parameters determining rolling effect performance are studied and a technique to measure the camera readout time per column is presented. Factors limiting the detectable frequency
range is explained based on the discussion of rolling shutter mechanism. Followed is the analysis of frequency detection reliability and resolution with Fourier spectrum analysis. After that, we present a background removal algorithm to recover the modulated information from background interference. In our algorithm, only one extra image is used and a one-time image alignment is required. Linear filter performs poorly in this case because the image with rolling shutter effect is a multiplication of modulated signal and the reflectance of the scene. Therefore we use background division rather than background subtraction to remove the background interference. The modulated illumination pattern is analyzed and the performance of background removal is evaluated by experiments. The experiment result shows that the interference is significantly compressed after background removal.

After the LED ID being detected successfully, an indoor positioning algorithm is proposed using multi-view imaging geometry with built-in smartphone inertial sensors. Due to the narrow field of view (FOV) of camera and the illumination placement of buildings, there is usually only one LED can be captured within camera FOV at the same time. In our proposed algorithm, only one detected LED with known position and one camera are required. And this system can still work when the VLC link is temporarily blocked or the LED temporarily moves out of the camera FOV. Another advantage of proposed algorithm is that no additional accessory will be needed except those sensors built inside the smartphone. In addition, the position is calculated at the receiver end without the knowledge of transmitter specifications, therefore there is no privacy concerns. The low-cost built in IMU sensor exhibits significant systematic errors, axes misalignment and noisy measurement. Even though a IMU calibration process can eliminate the systematic errors and axes
misalignment, the noise remains even after calibration and could magnificently degrade the system performance. To maximize a posterior of obtained measurements, Kalman filter is applied for the sensor fusion by combining the system kinematic prediction and new measurement. Due to the nonlinearity of this system, an extended Kalman filter is used to correct the system model. A simulation is conducted to verify the proposed system. The simulation data of accelerometer and gyroscope are generated based on real world measurement from accelerometer and gyroscope. The simulation results demonstrate that the position error and orientation error are well bounded in the $3\sigma$ bounds and the maximum position error observed during the 2 minutes simulation is 0.1941 m over 50 averaged Monte Carlo trials.
# Contents

<table>
<thead>
<tr>
<th>List of Figures</th>
<th>xii</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Introduction</td>
<td></td>
</tr>
<tr>
<td>1.1 Background of Indoor Positioning Systems</td>
<td>1</td>
</tr>
<tr>
<td>1.2 Motivation</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Objective</td>
<td>5</td>
</tr>
<tr>
<td>1.4 Contributions</td>
<td>6</td>
</tr>
<tr>
<td>1.5 Organization</td>
<td>7</td>
</tr>
<tr>
<td>2 Related work</td>
<td>9</td>
</tr>
<tr>
<td>2.1 Optical camera communication link characteristics</td>
<td>9</td>
</tr>
<tr>
<td>2.1.1 Calculating the pixel value from raw output</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Calculating the pixel value from the luminous exposure</td>
<td>10</td>
</tr>
<tr>
<td>2.2 VLC with rolling shutter camera</td>
<td>11</td>
</tr>
<tr>
<td>2.3 LED ID encoding and decoding</td>
<td>13</td>
</tr>
<tr>
<td>2.4 Optical ranging and communication</td>
<td>13</td>
</tr>
<tr>
<td>2.4.1 Principle of the ranging and communication system</td>
<td>14</td>
</tr>
<tr>
<td>2.4.2 Experiment Results</td>
<td>15</td>
</tr>
<tr>
<td>2.5 VLC based positioning algorithms</td>
<td>18</td>
</tr>
<tr>
<td>2.5.1 Fundamentals of VLC positioning</td>
<td>18</td>
</tr>
<tr>
<td>2.5.2 VLC based positioning algorithms</td>
<td>19</td>
</tr>
<tr>
<td>3 VLC using rolling shutter camera</td>
<td>28</td>
</tr>
<tr>
<td>3.1 Rolling shutter characteristics analysis</td>
<td>28</td>
</tr>
<tr>
<td>3.1.1 VLC with rolling shutter</td>
<td>29</td>
</tr>
<tr>
<td>3.1.2 Rolling shutter characteristics</td>
<td>30</td>
</tr>
<tr>
<td>3.2 Frequency limits</td>
<td>32</td>
</tr>
<tr>
<td>3.2.1 Frequency lower bound of optical tags</td>
<td>32</td>
</tr>
<tr>
<td>3.2.2 Frequency upper bound of optical tags</td>
<td>34</td>
</tr>
<tr>
<td>3.3 Frequency detection</td>
<td>37</td>
</tr>
<tr>
<td>3.3.1 Reliability of frequency detection</td>
<td>38</td>
</tr>
<tr>
<td>3.4 resolution of frequency detection</td>
<td>40</td>
</tr>
</tbody>
</table>
3.5 Background subtraction ................................................. 41

4 Multi-view Geometry Positioning with Extended Kalman Filter 50
4.1 Smartphone Sensor Overview ........................................ 50
  4.1.1 Motion sensors .................................................. 51
  4.1.2 Position sensors ................................................ 52
  4.1.3 Environmental sensors ......................................... 53
4.2 Camera Model ......................................................... 54
  4.2.1 Pinhole camera model .......................................... 54
  4.2.2 Lens distortions ................................................ 57
  4.2.3 Camera pose .................................................... 58
4.3 LED recognition on the image plane .................................. 61
4.4 Get orientation from accelerometer and magnetometer ............. 62
4.5 Positioning using multi-view geometry ............................... 64
  4.5.1 Kinematic equations ............................................. 67
  4.5.2 Simulation results of IMU propagation .......................... 72
4.6 Sensor fusion with extended Kalman filter ........................... 75
  4.6.1 Introduction to Kalman filter ................................... 75
  4.6.2 Implementation of extended Kalman filter ..................... 83
  4.6.3 Simulation with extended Kalman filter ....................... 85

5 Conclusion and Future Work 89
5.1 Conclusion .......................................................... 89
5.2 Future work ........................................................ 91

Bibliography 93
List of Figures

1.1 Photos taken from iPhone 6 front facing camera. The phone is held at a normal use height and only one LED light can be captured due to the narrow FOV. .......................................................... 5
1.2 System architecture .......................................................... 7

2.1 System block diagram ....................................................... 15
2.2 Experiment block diagram .................................................. 16
2.3 Position estimation using proximity. The workspace are divided into different area due to the occurrence of different signals .............................................. 21
2.4 Position are determined by comparing the measurements of RSS and AOA to database obtained in learning stage. ........................................... 22
2.5 Positioning using received signal strength (RSS) from different LEDs. .... 24
2.6 Positioning algorithm based on time of arrival (TOA). .................... 25
2.7 Position calculated by angle of arrival (AOA) from multiple LEDs. ...... 26
2.8 Two LED projection within one camera view. ............................ 27
2.9 Two camera view imaging geometry. .................................... 27

3.1 Rolling shutter based VLC receiving mechanism: the bright and dark rows of the camera sensor indicate the LED on-off status. Transition bands appear because of the long exposure time overlap with both LED on-time and off time. .......................................................... 29
3.2 Transmitted analog frequency V.S peak digital frequency. The measured $\tau_{\text{readout}} = 12.65 \, \mu s$ with iPhone 6 at 1/1000 second exposure time and 3264$ \times $2448 picture resolution. The red line is a linear fitting of measurement. .... 31
3.3 Photos taken at different shutter speed and ISO. The position of bands shifts in different frames because the modulation frequency is not an exact multiple of the video frame rate. ........................................... 33
3.4 (a) Received 800 Hz signal with nonuniform background and ambient. (b) The FFT amplitude of pure background scene. (c) The FFT amplitude of modulated signal combined with background scene. (d) The FFT amplitude of received signal after background removal operation. ............. 35
3.5 Frequency response of camera ............................................ 36
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.6</td>
<td>Effect of exposure time to the received signal</td>
<td>36</td>
</tr>
<tr>
<td>3.7</td>
<td>(a) nonuniform reflectance of the background; (b) modulated illumination;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(c) captured image with rolling shutter effect; (d) average of a sequence of</td>
<td></td>
</tr>
<tr>
<td></td>
<td>frames of images with rolling shutter effect; (e) directly subtract Fig. 3.7d</td>
<td></td>
</tr>
<tr>
<td></td>
<td>from Fig. 3.7c; (f) divide Fig. 3.7c by Fig. 3.7a.</td>
<td>43</td>
</tr>
<tr>
<td>3.8</td>
<td>(a) Pure background image captured with longer exposure time. (b) Signal</td>
<td></td>
</tr>
<tr>
<td></td>
<td>image with rolling shutter effect.</td>
<td>47</td>
</tr>
<tr>
<td>3.9</td>
<td>Fourier spectrum analysis</td>
<td>48</td>
</tr>
<tr>
<td>4.1</td>
<td>Coordinate system used by motion sensors, image is copied from Android</td>
<td>51</td>
</tr>
<tr>
<td></td>
<td>developer website [1].</td>
<td></td>
</tr>
<tr>
<td>4.2</td>
<td>Coordinate system used by motion sensors, image is copied from Android</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>developer website [1].</td>
<td></td>
</tr>
<tr>
<td>4.3</td>
<td>Pinhole camera model: a pinhole aperture projects only those rays intersected</td>
<td></td>
</tr>
<tr>
<td></td>
<td>with optical axis to image plane.</td>
<td>55</td>
</tr>
<tr>
<td>4.4</td>
<td>Rearrangement of image plane and aperture plane. The point P is project</td>
<td></td>
</tr>
<tr>
<td></td>
<td>to the image plane p, the negative sign is eliminated by this rearrangement.</td>
<td>56</td>
</tr>
<tr>
<td>4.5</td>
<td>Radial distortion: rays at the edge of an image are bent more than those</td>
<td></td>
</tr>
<tr>
<td></td>
<td>close to the center of the image, known as barrel distortion or fish eye</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>distortion.</td>
<td></td>
</tr>
<tr>
<td>4.6</td>
<td>Tangential distortion caused by the misalignment of the lens and image sensor</td>
<td></td>
</tr>
<tr>
<td></td>
<td>during the assembling process of camera.</td>
<td>58</td>
</tr>
<tr>
<td>4.7</td>
<td>Converting point in global coordinates system to camera coordinates system:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>the point P in the global frame is projected to the image plane as point p;</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>the point p is related to point P by applying a rotation matrix and a</td>
<td></td>
</tr>
<tr>
<td></td>
<td>translation vector.</td>
<td></td>
</tr>
<tr>
<td>4.8</td>
<td>Comparison of position propagation using IMU data and position of ground</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td>truth.</td>
<td></td>
</tr>
<tr>
<td>4.9</td>
<td>Comparison of quaternion propagation using IMU data and quaternion of</td>
<td>74</td>
</tr>
<tr>
<td></td>
<td>ground truth.</td>
<td></td>
</tr>
<tr>
<td>4.10</td>
<td>Comparison of position propagation using IMU data and position of ground</td>
<td>76</td>
</tr>
<tr>
<td></td>
<td>truth under noise condition.</td>
<td></td>
</tr>
<tr>
<td>4.11</td>
<td>Comparison of quaternion propagation using IMU data and quaternion of</td>
<td>77</td>
</tr>
<tr>
<td></td>
<td>ground truth under noise condition.</td>
<td></td>
</tr>
<tr>
<td>4.12</td>
<td>Position error propagation.</td>
<td>78</td>
</tr>
<tr>
<td>4.13</td>
<td>Rotation angle error propagation.</td>
<td>79</td>
</tr>
<tr>
<td>4.14</td>
<td>Position propagation corrected by extended Kalman filter (EKF).</td>
<td>85</td>
</tr>
<tr>
<td>4.15</td>
<td>Orientation propagation corrected by extended Kalman filter (EKF).</td>
<td>86</td>
</tr>
<tr>
<td>4.16</td>
<td>Position error and the $3\sigma$ bounds.</td>
<td>87</td>
</tr>
<tr>
<td>4.17</td>
<td>Orientation error and the $3\sigma$ bounds.</td>
<td>88</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 Background of Indoor Positioning Systems

The Global Positioning System (GPS) is the most well-known system when it comes to positioning and tracking. The GPS is able to provide location and time information under all weather conditions, whenever there is a line-of-sight receiving link from four or more GPS satellites [6]. However, the GPS signal suffers from great attenuation and multi-path effects when the GPS signal penetrates buildings and constructions, making it fail to acquire position information in indoor environment. Indoor positioning systems have utility in a variety of environments. For example, it provides guidance through complex buildings like airports and train stations; it can be used for location-based advertisements and marketing in the retail industry; it makes asset tracking possible in warehouse and hospital. In order to complement GPS for indoor navigation, techniques based on receiving radio frequency (RF) power intensity employing cellular [8, 32], WiFi [53, 25], Bluetooth [48, 57] as well as RF identification (RFID) [54, 36] have been developed for indoor po-
sitioning services. Most of these systems measure the receiving signal strength or the propagation delay from multiple transmitters to estimate the distance between receivers and transmitters. The position of receiver is determined using geometrical method. The precision significantly depends on the mobile device and indoor environment. Systems using non-audible acoustic signal to localize the smart phone user without additional hardware is also exploited [16, 15].

In recent years, indoor positioning based on Visible Light Communication (VLC) has become an attractive technique as a promising solution to indoor positioning and tracking. Position estimation using VLC refers to obtaining location information with respect to a set of positions of reference LED within a defined space. The line-of-sight property enables the receiver to sense the relative position and rotation with respect to each detected LED. The position of receiver can be calculated based on the position of each LED transmitter. VLC based positioning has several advantages over previously mentioned RF based techniques. First, the line-of-sight propagation property gives higher precision when utilizing imaging geometry to calculate position. This will greatly reduce the environmental influences and the uncertainty caused by different types of smartphones and antenna orientation. Another major benefit of VLC based positioning is that the multi-path effects is negligible compared to RF based counterpart. In addition, the visible light positioning can cover the scenarios where RF is restricted, for example, in hospital. Besides these advantages, the VLC based positioning systems utilize existing illumination systems which means only a few modifications will be needed to current infrastructure. These unique properties make camera based visible light positioning a promising indoor positioning technique.
In general, there are two different types of optical receiver for VLC: photo diode and camera. Photo diode promises a high data-rate and low-cost optical wireless link, while the camera based VLC, though more expensive, has the advantage of no additional hardware requirement when mobile devices with cameras are employed as receivers. In our application, the purpose of using VLC in indoor positioning is to transmit LED IDs, so a smartphone without additional hardware is preferable here for its ubiquity. In traditional optical camera communication, the LED data is modulated using an on-off keying (OOK) scheme. Multiple camera frames are recorded to decode the on and off status of LED. Even though the resolutions of image sensors have been pushed over 10 megapixels, the frame rate stays relative low for a low-cost camera. The data rate is therefore limited by the camera frame rate, which is usually up to 120 frames per second. More importantly, the low frame rate camera limits not only the data rate but also the LED modulation rate, which will introduce severe undesirable flickering effects to human eyes. In addition, it is not so straightforward to recover the LED ID from a sequence of images. This is because: 1) The clutter and noise increase the uncertainty of the LED status; 2) the movement of mobile user changes the projection of LED on the image sensor, which makes the detection more difficult, especially in those frames when the LED is off. To increase the data rate, LED array are used to transmit LED pattern, which increases the transmitted bits per frame [37]. This approach needs complex image processing techniques to decode the information and the flickering effect still remains.

Recently, the rolling shutter effect of Complementary Metal-Oxide-Semiconductor (CMOS) cameras has been exploited to allow a much higher data rate than the camera
frame rate. Unlike global shutter, which exposes camera to the entire scene at the same
time, each row of pixel within a rolling shutter camera starts integrating sequentially [14].
The effective rolling shutter mechanism potentially increases the LED modulation frequency,
which in turn eliminates the LED flickering regardless of the low frame rate. Also, a single
frame of image is enough to extract the LED ID instead of using a sequence of images.

1.2 Motivation

First, for most work related to VLC employ rolling shutter mechanism, they just
examine those cases with close range to modulated light and uniform reflecting surface under
no ambient light, which is usually not the case in practice. A comprehensive investigation of
the rolling shutter receiving mechanism is needed. The effects of the camera parameters to
the receiving performance should also be studied. More important, an LED ID extracting
algorithm that can deal with ambient light and nonuniform reflectance background should
be proposed for practical application.

Secondly, for VLC based indoor positioning algorithm, the PD based receiving
algorithm is not available on smartphones since it requires additional hardware. The camera
based positioning algorithm require at least two LEDs with position known. However,
multiple LEDs are not always feasible due to the narrow field-of-view of camera and the
illumination placement inside the buildings. Usually, only one LED or none can be captured
within the camera FOV (Fig. 1.1). Furthermore, the positioning result of most current
approach is rather binary: either it knows its position, or not. It is not capable of these
situations when the VLC link is temporarily blocked or the LED temporarily moves out
from the camera FOV. So the proposed positioning algorithm should work with one detected LED and gives best estimation when the only LED is temporarily unavailable.

1.3 Objective

The objectives of this research is to propose an indoor VLC positioning system with high accuracy that works with nonuniform reflectance as well as ambient light using only one detected LED. Users with smartphone can achieve indoor navigation and positioning without extra hardware. Easy modification is just needed at the existing illumination systems. For detection purpose, signal processing and wireless communication approach will be used to extract the signal information. In positioning stage, the image processing and image geometry will be applied for accurate locating purpose since the optical beam is highly directional. The positioning algorithm can work with only one detected LED and can seamlessly switch between one LED mode and multi-LED mode without dependency.
on the knowledge of transmitter shape, radiation and transmitted power. The position is calculated at the receiver end without extra hardware, thus there is no privacy concerns.

The main structure of the indoor visible light positioning system is illustrated in Fig. 1.2. Each illumination LED is modulated at a unique frequency used as optical tag. Square waveforms with different frequencies generated by a low complexity LED driver can be used to drive the LED to broadcast the optical beacon. The smartphone camera will extract the beacon information from LED directly or from a reflected surface by analyzing rolling shutter effect. The modulated light will result in bright and dark bands with different width in the image. The image is being further processed to extract the frequency component corresponding to different LEDs. With the pre-known location of different LEDs using a look-up table offline or from a server online, the user position is calculated through an imaging geometry method with information provided by other sensors built inside the smartphone.

1.4 Contributions

The main contributions in this dissertation are:

- A comprehensive study of rolling shutter effects. A detailed analysis of the effects of camera parameters to the rolling shutter receiving performance.

- Theoretical analysis of the lower bound, upper bound of the frequency of transmitted signal as well as the minimum frequency interval between them, thus the system capacity can be derived.
• A background removal algorithm is proposed for rolling shutter based VLC receiving, which can significantly reduce the background interference using only one extra image.

• A positioning algorithm using multi-view geometry with single camera, which only needs one detected LED with known position for narrow FOV camera. An extended Kalman filter is applied to further refined the measurements. This system still give position estimation when the VLC link is temporarily blocked or the LED temporarily moves out from the camera FOV.

1.5 Organization

This dissertation is organized as follows: Chapter 2 introduces the work have been done related to rolling shutter VLC and indoor VLC based positioning algorithm.
Chapter 3 gives a comprehensive study of rolling shutter effects. The factors affecting the rolling shutter VLC receiving is discussed. The system capacity is also given in this chapter. At the end of this chapter, a background removal algorithm is introduced to remove the background interference. In chapter 4, we propose a multi-view geometry VLC positioning algorithm which can work with only one LED. The implementation of extended Kalman filter is given in this chapter. Finally, chapter 5 concludes the work of this dissertation and outlines the remaining problems needed to be tackled and potential solutions and approaches to them.
Chapter 2

Related work

In this chapter we first review existing research work related to VLC based indoor positioning using rolling shutter, followed by a comparison of different methods to encode and decode LED ID. Finally, the indoor positioning algorithms based on VLC are reviewed and compared from fundamental papers to the state-of-the-art in this area.

2.1 Optical camera communication link characteristics

In the paper [10], the authors developed a model to calculate the pixel value from the luminous exposure in order to analyzed the performance of optical camera communication link. In a camera receiving process, the light passes through the lens and reaches on the image sensor. The amount of light falling on per unit area of the image sensor, is determined by the lens aperture, exposure time of the camera as well as a specific luminous exposure of the object. After that, the optical energy is converted to a voltage and being amplified. The international standard organization (ISO) speed decides how much of the
voltage is being amplified during a amplification process. Then the voltage will be digitalized through an analog-to-digital converter and this digital value is the raw output of the image sensor. Finally, the digital value will be quantized to a pixel value with a logarithmic operation.

2.1.1 Calculating the pixel value from raw output

The image sensor raw output is proportional to the luminous exposure. However, a gamma encoding operation will redistribute the raw output values from the image sensor, since the response of human eyes to the light intensity follows a logarithmic sense. The gamma encoding is a logarithmic operation that mimics the process of how human eyes perceive light:

\[ V_{\text{pixel}} = 255 \times \left( \frac{\text{raw}}{\text{raw}_{\text{max}}} \right)^{1/\gamma} \]  

with \( V_{\text{pixel}} \) being the pixel value, \( \text{raw} \) and \( \text{raw}_{\text{max}} \) representing raw output value and its maximum, respectively. \( \gamma \) is the logarithmic index, the standard value of which is 2.2.

2.1.2 Calculating the pixel value from the luminous exposure

The luminous exposure \( H \) is given by:

\[ H = \frac{qL_{\text{ext}}t}{N^2} \]  

(2.2)
where $L_v$ is the object luminance (cd/m$^2$), $N$ is the lens relative aperture, $t$ is the exposure time (second) and the factor $q$ is determined by:

$$ q = \frac{\pi}{4} T v(\theta) \cos^4(\theta) \quad (2.3) $$

with $T$ as the transmittance of the lens, $v(\theta)$ as the vignetting factor and $\theta$ being the relative angel to the optical axis of the lens.

Since the raw output is proportional to the luminous exposure, we have:

$$ \text{raw} = c \times H \quad (2.4) $$

where $c$ is a constant. Combining Eq. 2.1, Eq. 2.2 and Eq. 2.4, we have:

$$ V_{\text{pixel}} = 255 \times \left( \frac{c q L_v t}{\text{raw}_{\text{max}} N^2} \right)^{1/\gamma} \quad (2.5) $$

Eq. 2.5 is very useful in the performance analysis of the system, like signal quality and noise impact.

### 2.2 VLC with rolling shutter camera

In year 2012, ByteLight filed a patent using a rolling shutter camera to detect alternating stripes in the image caused by the modulation of LED light. The width of the alternating stripes, corresponding to unique LED ID, is measured to recover information encoded in the frequency of light pulses. However, very little detail is given in this patent.
In the same year, the authors in [7] explored the possibility of using rolling shutter effect to decode OOK modulated data from LEDs. The information bits are encoded using Manchester coding following a inserted header for synchronization purpose. Image processing techniques are applied to convert these bright and dark bands into a binary array form. A baud rate up to 3.1 kBd is achieved using a camera with 640 × 480 pixels at a 20 frames per seconds (FPS) rate. However, this requires a close distance to the LED and a nearly uniform reflecting surface. Furthermore, flicker could be observed due to the inserted header pattern at low baud rate. This paper established the concept, demonstrating that the rolling shutter can be used for VLC.

After that, lots of papers related to this topic have been published. [41] presents an approach that is able to demodulate different IDs from different LEDs operating at different frequencies simultaneously in a collision domain. They also built a prototype to evaluate the real-world performance of this scheme. In [23], the authors study the line-of-sight light-to-camera link channel characteristics, and then theoretically and experimentally examine the demodulation Error Probability as well as throughput performance.

One of the greatest challenges in practice is to extract the LED ID from the image polluted by background and ambient light. De-focusing the image is a simple way to deal with the background influence. This operation acts like a low-pass filter to remove the high frequency components in the background while retaining the signal spectrum [41]. ByteLight proposed a background subtraction algorithm to remove the contribution of the background scene [11]. The bright and dark bands may appear at different location on different frames when the modulation frequency is not multiple of the frame rate. The
average of multiple frames will cancel out the effects of modulated illumination light and
give a pure background scene without the illumination modulation. Due to the movement
of mobile user, motion compensation is necessary before applying frame average.

2.3 LED ID encoding and decoding

Either a unique sequence of bits or a unique frequency can be assigned to LED ID.
These two different LED ID solutions corresponds to two different decoding mechanisms.
When a bit sequence is used to represent an LED ID, a light intensity threshold is usually
applied to decode ID information. Pixel intensity above threshold will be regraded as 1, and
0 otherwise. Polynomial fitting could be used in a non-homogeneous image to normalize
the data sequence [7]. Image processing techniques are also used to enhance the VLC
signal and decrease the impact of unbalanced illumination [33, 5]. For frequency based
identification, Fast Fourier transform (FFT) is usually employed to analyze the frequency
spectrum [41, 11]. In addition, the band width and the number of the bright and dark bands
are also used to measure the modulation frequency [20, 22]. Besides, Bytelight proposed an
alternative algorithms using the phase variance to detect modulation frequency [11].

2.4 Optical ranging and communication

In the work [29], we developed a ranging and communication system for the vehic-
ular safety applications. It shows how VLC can be used to acquire information and location
simultaneously. The LED array and Photo Detector (PD) array serve as the transceiver.
The imaging technique is applied to the receiver, which can detect the angle of incident
light and imaged different rays to corresponding PDs. Benefiting from this structure, this system is able to detect distance as well as location, and work between multiple users simultaneously. While keeping a low-cost and compact size, it can still provide high lateral resolution, high frame rate and interference rejection performance.

2.4.1 Principle of the ranging and communication system

The original idea of ranging and communication using spread spectrum (SS) waveform was proposed by [34]. However, they just used the laser at the transmitter end and a single PD at the receiver end, which will bring the problems: costly, bulky, requiring fine scanning mechanism and complicated tracking algorithm. Furthermore, it can only measure the distance instead of location and only support point-to-point operation rather than between multiple vehicles.

Fig. 2.1 demonstrates the block diagram of the proposed ranging and communication system. First, vehicle No.1 transmits a pulsed spread spectrum signature waveform $s_1(t)$ to vehicle No.2. Then, vehicle No.2 captures the signal via an image lens and renews its neighbor list table (Table 2.1). After this, vehicle No.2 multiplies information bits (SS modulation) to the signature waveform and transmits back to vehicle No.1. Vehicle No.1 acquires the distance measurement by sensing the time-of-flight (TOF) between the emitted light and received light and then extract the information bits.

This system can work under two different modes. Even though the vehicle No.2 is not equipped with this system, vehicle No.1 can still acquire the distance via the reflecting signal.
In this process, vehicle No.1 can not only determine the distance but also the location of vehicle No.2, because imaging method can provide the angle information of incoming light. And if two or more users receive the ranging waveform then respond to it simultaneously, their signal will be imaged to different areas of the PD array. The interference between different users can be mitigated significantly. The reflecting signal power falling into the corresponding PD, which mainly senses the active responding signal, is also limited in the case that there is only one single user responds.

2.4.2 Experiment Results

Fig. 2.2 illustrates the experiment block diagram. On the left end of the diagram, an M sequence was generated by an arbitrary waveform generator (AWG) to drive LED with the On-off-key modulation. The length of the M sequence is 31 and the chip width is

![Figure 2.1: System block diagram](image)

Table 2.1: Neighbor list table of vehicle No. 2

<table>
<thead>
<tr>
<th>Neighbor</th>
<th>Position</th>
<th>Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>( s_1(t) )</td>
<td>((x_1, y_1))</td>
<td>(d_1)</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
<tr>
<td>( s_k(t) )</td>
<td>((x_k, y_k))</td>
<td>(d_k)</td>
</tr>
<tr>
<td>( \ldots )</td>
<td>( \ldots )</td>
<td>( \ldots )</td>
</tr>
</tbody>
</table>
A single PD was applied to capture the active responding signal with an imaging lens to collect the energy. The captured signal was recorded by an oscilloscope and then processed by MATLAB offline. On the right end, another single PD was used to receive the ranging signal. The signal was then transmitted back to the left end after being amplified. A customized PD array was not available so we used a single PD instead. But we can still prove the imaging method and PD array structure are beneficial later in this experiment.

The experiment was completed indoor with all office artificial lights on as the ambient light noise. The two terminals were separated at a distance of 5 m, 10 m, 15 m, 20 m, 25 m and 30 m. At each distance, ten measurements were conducted. And after 20 m, a huge glass about 25 m far behind the receiver was applied as the reflection interference. So the data acquired after 20 m contains the reflection noise signal. Table 2.2 gives the measured average time of flight at each distance which is the true time of flight plus circuit delay.
Table 2.3: Measured distance difference results

<table>
<thead>
<tr>
<th>Distance Difference</th>
<th>Measured Distance</th>
<th>Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5m – 10m</td>
<td>5.331m</td>
<td>6.62%</td>
</tr>
<tr>
<td>10m – 15m</td>
<td>5.058m</td>
<td>1.16%</td>
</tr>
<tr>
<td>15m – 20m</td>
<td>5.229m</td>
<td>4.58%</td>
</tr>
<tr>
<td>20m – 25m</td>
<td>4.209m</td>
<td>15.82%</td>
</tr>
<tr>
<td>25m – 30m</td>
<td>5.160m</td>
<td>3.2%</td>
</tr>
</tbody>
</table>

We can’t directly use these data to calculate the absolute distance because the measured time contains circuit delay. So we calculate the relative difference between two different distances instead of absolute distance. Here we assume that the circuit delay remains the same at different distance. Table 2.3 shows the obtained results and deviations.

From the table, we can find there is a large deviation (more than 15%) between 20m-25m, this is because the glass brings in reflecting noise after 20m. The reflecting noise was canceled out again, so the deviation decrease to 3.2% between distance 25m – 30m. Except this point, we can see that the proposed system is robust to the ambient light noise. The PD was set in the focal plane of the lens. The lens before the PD focus the light into a small light spot. The PD couldn’t detect the transmitted signal once the focused light spot was not imaged to the PD sensor area. So if you use another LED to simulate the different user interference, it will be imaged to outside of the PD sensor area and won’t affect the received signal. But if we use a PD array as the receiver, it will be imaged to another PD and demodulated correctly. This illustrates that our proposed system is also robust to the reflection interference and user interference.
2.5 VLC based positioning algorithms

2.5.1 Fundamentals of VLC positioning

VLC based positioning is a technique of using VLC signal transmitted from LED (or lamp) referring to the position of the LED to locate a mobile user. Usually the physical coordinates of the LED is transmitted via the VLC link while a PD or camera is usually used for receiving the optical signal. In the case where the LED ID is transmitted, the mobile device needs either an off-line pre-installed map or real time access to the data server which stores the position of the corresponding LED, in order to map the LED ID to their physical coordinates. Then the VLC based positioning become the problem to determine the relative position between the LED and the mobile device, taking the advantage of the line-of-sight optical link.

For a room-size positioning application, a simple PD based receiver with wide field of view (FOV) can be applied to extract the optical positioning information with simple signal processing. These receivers have the advantage of being able to detecting optical signal with relatively high frequencies. This is crucial in order to avoid flickering effects, especially in office environment. Wide FOV is another advantage of PD based receiver, which enables it to receiver signals from multiple transmitters simultaneously. However, PD based receiver can only estimate the time-of-arrival and the strength of the signal since it directly measures the optical power and converts it to electrical current proportional to the incident optical power. A angle-of-arrival (AOA) can only be roughly estimated due to the wide FOV. So PD based receiver can exploit the received signal strength, time-of-arrival and AOA of received signal to determine the location of mobile device. To use these
characteristics of signal, the information of transmitter must be known in advance, like transmitted power and FOV of LED. Assumptions such as Lambertian light source and uniform radiation are also made to simplify the system model.

To acquire the AOA information for accurate positioning, imaging geometry can be applied to receive the optical signal. A camera based VLC systems is preferred to determine the signal AOA. The position of the LED in the image decides the relative location between the camera and LED. Unfortunately, the low frame rate of commercial camera limits the frequency of detectable signal, which will bring in the undesirable flickering effects. Good thing is that the use of proposed rolling shutter receiving technique can increase the frequency of optical signal and avoid flickering effects. Another drawback of camera is that its narrow FOV might not capture enough numbers of LEDs to do the triangulation positioning.

2.5.2 VLC based positioning algorithms

In this section, we briefly overview the indoor VLC positioning methods with different positioning algorithms and different types of receivers. Literature [9] gives an in-depth review of VLC positioning systems based on more than 100 papers ranging from pioneering papers to the state-of-the-art in this field. Here we only give a summary of different positioning algorithm, please refer to [9, 13] for more details. The algorithms in VLC based algorithms can be broadly classified into four categories: proximity, fingerprinting, triangulation and computer vision, regardless of the receiver types.
Proximity

Proximity is the simplest VLC based locating algorithm which can only give an approximated position. The location approximation is determined by the appearance of the optical signal (either ID or absolute coordinates). The proximity algorithm is illustrated in Fig. 2.3, the workspace are divided into different area due to the occurrence of different signals. The position is determined by the signals sensed at the receiver end. The accuracy is decided by both FOVs of transmitter and receiver. However, these positioning systems can be very useful in applications with room-size accuracy like assets tracking [2], due to its simple implementation and low cost. In such systems, RF techniques such as bluetooth or WiFi are usually required to access the database server to poll the location information corresponding to the LED ID.

In such systems, PD is commonly used as the receiver for simplicity. Modulated frequency can be relative high to avoid flickering effects. The resolution of systems utilizing proximity algorithms is several meters. Additional information provided by RF signal link and auxiliary sensors such as 6-axis sensor and geomagnetic sensor can be combined to improve the position accuracy [42, 35]. The system performance could be improved by detecting multiple LEDs at the same time. Multiple LEDs are applied to create overlapped regions. Each region is distinguished by the appearance and absence of specific LEDs [44]. Increasing number of LEDs and narrowing down the receiver FOV can increase the system performance.
Figure 2.3: Position estimation using proximity. The workspace are divided into different area due to the occurrence of different signals

**Fingerprinting**

Fingerprinting based positioning algorithm estimates the position by matching the real-time measurement with the pre-trained position-related data. Fingerprinting based algorithm is more like a machine leaning based algorithm. An off-line training stage is necessary before on-line estimation. The working scene is divided into several different areas by the features of position related information gathered from local environment. In the on-line positioning stage, the position is estimated by comparing the measurement and the stored information. The received signal strength (RSS) is most commonly used for fingerprinting. However, other features of signal, such as time difference of arrival (TDOA), angle of arrival (AOA), correlation between different transmitters can be modeled to extract
the location pattern. Fig. 2.4 shows that the position is further refined by comparing the measurements of RSS and AOA to database obtained in learning stage. The patterns of these features can be recognized by applying Bayes analysis [56], analyzing signal correlation [18], k-nearest neighbors and neural networks[31, 46].

Figure 2.4: Position are determined by comparing the measurements of RSS and AOA to database obtained in learning stage.

The complexity of this algorithm lies in creating the feature database from environment, and the system performance depends on how different these features lie in different areas.
**Triangulation**

Triangulation is the most popular technique in ranging and positioning application. In triangulation algorithm, the RSS, TDOA and AOA information will be involved in the calculation to determine the receiver location.

Received Signal strength (RSS): The receiver evaluates distances to multiple transmitters by measuring the received signal strength. After that, the position of the receiver is determined using triangulation. In order to derive the distance to LED based on the received signal strength, the parameters of both transmitter and receiver, as well as the optical link model must be pre-known. In order to more accurately calculate the received signal strength, motion sensor (accelerometer and magnetic field sensor) can be used to obtain the orientation information of receiver [17]. The system performance is improved by taking the irradiance angle and incidence angle of the received signal into consideration. In order to separate independent LED ID from received mixed signal, multiplexing techniques like frequency division or optical orthogonal code should be used in this algorithm [50].

Time of Arrival (TOA): Similar to RSS, TOA derives the distances to different transmitter from the signal arrival time of signal. The distance is calculated by directly multiplying light velocity to the propagation delay. Then the receiver position is determined via the distance to different the transmitters. This kind of system requires a synchronization between transmitter and receiver. Due to the fast light speed and short propagation distance, it requires high resolution and high accuracy system clock on both transmitter and receiver sides. In order to estimate a two-dimension coordinates, at lest three transmitter need to be used at the same time. And a least-square algorithm is applied to find
the optimal estimation of receiver position [47].

Time Difference of Arrival (TDOA): Instead of measuring the propagation delay of different links between transmitters and receiver, TDOA measures the time difference of signal arrival between different communication links [21]. There is no synchronization requirement between transmitter and receiver. However, it does require synchronization between all transmitters in order to obtain accurate time difference. Same as TOA algorithm, at least three LEDs are required to estimate the receiver position.

Angle of Arrival (AOA): AOA method evaluates the angles of signal arrival from different LEDs to determine the receiver position. The receiver position is constrained by the overlapped common volume of FOVs of corresponding LEDs. And there is no
synchronization between transmitter and receiver or between transmitters. Compared to RF counterpart, optical systems take the advantage of light line-of-sight propagation. To measure the angle of signal arrival, usually more than one PD is used, for example, PD array [24]. The received power of PD array depends on the distance between transmitter and receiver, radiance angle and the incidence angle. The distance to each PD and irradiance angle of incoming light are assumed to be the same. So the difference of received power is caused by the difference of incidence angle only. Then the angle of arrival is calculated based on the difference of received power on each PD.
When it comes to the computer vision based algorithm, it is mainly focusing on the camera receiving method. The computer vision analyzes imaging geometry to determine the receiver 3D position using 2D position of LED on the image sensor. To analyze the imaging geometry, the intrinsic parameters (like the focal length) and the orientation (or pose) of the camera must be known. If the pose of the camera is unknown, three LEDs should be captured simultaneously to determine the receiver location. When one camera is used to capture the LEDs, at least two LEDs are needed to obtain the camera location under the condition the orientation of camera is known. Since the depth information is losing when transforming a 3D point to a 2D point, at least two LEDs with pre-known location should be used to calculate the camera position (Fig. 2.8). Rather than using multiple LEDs, multiple cameras can be used to capture the scene (Fig. 2.9). The relative position and orientation between cameras must be known in order to compute the position.
of receiver.

Figure 2.8: Two LED projection within one camera view.

Figure 2.9: Two camera view imaging geometry.
Chapter 3

VLC using rolling shutter camera

This chapter gives a concise outline of the work that have been carried out so far and of the progress toward the aims of the project. The obtained preliminary results is presented to support this proposal. For now, the signal detection part, including the rolling shutter characteristics and detection technique have been studied.

3.1 Rolling shutter characteristics analysis

The global shutter and the rolling shutter are two different kinds of electronic shutters for CMOS image sensors. Both of them are electronic shutters and require no mechanical shutter, but they have different exposure sequences [40]. In global shutter mode, each row of pixels is exposed simultaneously and its exposure process ends at the same time point. After the exposure of each pixel is completed, the analog-to-digital converter (ADC) traverses through the sensor and digitizes each row of pixels separately. Next exposure will not begin until the digitization and sensor clear procedure are completed. The global shutter
Figure 3.1: Rolling shutter based VLC receiving mechanism: the bright and dark rows of the camera sensor indicate the LED on-off status. Transition bands appear because of the long exposure time overlap with both LED on-time and off time.

is especially beneficial when capturing a fast moving scene at the cost of sacrificing frame rate. The rolling shutter mechanism is proposed in order to fully exploit the frame rates. Instead of exposing each pixel simultaneously and waiting for a completely pixel readout, each individual row will begin next exposure as soon as the current row is read out. While this further maximizes the frame rates, the previous row readout transfers a delay to next row at the beginning of exposure since the readout of each row cannot overlap, as illustrated in Fig. 3.1. This results in a time shift at the beginning of each row with respect to the previous row and makes rolling shutter a sequential readout architecture.

3.1.1 VLC with rolling shutter

The unique character of rolling shutter makes the CMOS sensor act like a high speed sample-and-hold ADC. The only difference is that the CMOS has a large portion of overlap between two sampling points caused by the overlapped exposure of two adja-
cent rows. This effect is similar to the inter symbol interference caused by the multipath propagation. When the LED light is modulated by a square waveform, those rows whose exposure time overlaps the LED on-time will accumulate more energy, while those overlaps the LED off-time have lower integrated energy. Consequently, bright and dark bands on the image will be observed (Fig. 3.1). The numbers of bright and dark rows are proportional to the LED on-time and off-time respectively. Due to this property, pulse-width modulation (PWM) can be utilized to the modulate LED light. On-off keying can also be used since the bright band and dark band can represent bit 1 and bit 0. Besides, the frequency-shift keying modulation scheme, where different frequencies denote different information, can be adopted. The rolling shutter mechanism allow a much higher data rate compared to a camera frame based detection method which is limited by the frame rates. However, the exposure overlap could distort the original signal.

3.1.2 Rolling shutter characteristics

For rolling shutter based VLC system, the readout time of each row $\tau_{\text{readout}}$ is a key factor affecting the system performance. It can be seen from Fig. 3.1 that $\tau_{\text{readout}}$ is actually the sampling interval. For a CMOS sensor, the readout time mainly depends on the speed (clock frequency) of ADC as well as the number of column of the sensor. The $\tau_{\text{readout}}$ can be approximated as $\tau_{\text{readout}} \approx 1/f_{\text{ADC}} \cdot N_{\text{col}}$, where $f_{\text{ADC}}$ is the ADC speed and $N_{\text{col}}$ is the number of pixels in each row needed to be processed by ADC. However, the readout time is usually not provided by the camera manufacturer, nor the ADC speed of camera. $\tau_{\text{readout}}$ should be measured at the design stage to correctly detect the received signal frequency. In our experiment, an iPhone 6 is used as the receiver. The camera app
Figure 3.2: Transmitted analog frequency V.S peak digital frequency. The measured \( \tau_{\text{readout}} = 12.65 \, \mu s \) with iPhone 6 at 1/1000 second exposure time and 3264 × 2448 picture resolution. The red line is a linear fitting of measurement.

is VSCO, with 1/1000 second shutter and 3264 × 2448 resolution. To evaluate the readout time, we vary the transmitted frequency from 200 Hz to 2000 Hz and detect the peak point in received digital frequency spectrum. The relations between transmitted analog frequency and the peak digital frequency is plotted in Fig. 3.2. The approximately estimated readout time is \( \tau_{\text{readout}} = 12.65 \, \mu s \), which corresponds to a sampling frequency \( f_s = 79 \, \text{kHz} \).

The exposure time (shutter speed) is another important parameter to the system performance. A straightforward approach to decide whether an image will experience rolling shutter effect is that whether the exposure time is larger than frame rate \( \tau_{\text{frame}} \) or not. The frame rate is determined by the readout time \( \tau_{\text{frame}} = \tau_{\text{readout}} \cdot N_{\text{row}} \) [40], where \( N_{\text{row}} \) is the total number of rows of pixels on the CMOS sensors. If exposure is shorter than \( \tau_{\text{frame}} \), rolling shutter effect is expected to occur. As can be seen from Fig. 3.1, the exposure time also bring in transition bands between bright bands and dark bands, whose intensity is
higher than dark bands but lower than bright bands. These transition bands are caused by the long exposure time. For those rows whose exposure time overlap with both LED on-time and off-time, they will accumulate less optical energy than those totally covered by the LED on-time but more than those exposed during the LED off-time. Short exposure time not only reduces the width of transition bands but also better reflects the true duty cycle of the received signal so fast shutter speed is preferred. However, short exposure also decrease the optical energy reaching on the sensor, resulting an under exposed image. The camera exposure time and ISO value should be optimized based on the application scenario. Fig. 3.3 illustrates how different exposure time affects received signal. We can see the interval between two adjacent bright bands is the same regardless of exposure time. This indicates the exposure time will not change the pattern of received signal but the proportion of bright and dark bands. Short exposure time produces images with sharp bands but it could lead to an under exposed dark image. Increasing ISO can compensate the signal contrast at the cost of increasing the noise.

3.2 Frequency limits

3.2.1 Frequency lower bound of optical tags

First, we examine the lower bound of detectable frequency $f_{\text{min}}$. To successfully observe at least one period of signal, the signal period should be at most one frame time $\tau_{\text{frame}}$. Therefore, the lower bound $f_{\text{min}}$ should be at least the order of $1/\tau_{\text{frame}}$. Another factor needed to be taken into consideration is the flicker effect when examining the fre-
Figure 3.3: Photos taken at different shutter speed and ISO. The position of bands shifts in different frames because the modulation frequency is not an exact multiple of the video frame rate.
quency lower bound. Normally, the light flicker effect can be negligible once the modulation frequency exceeds 60 Hz. But in rare instances the flicker can be sensed at 100 – 110 Hz. So the LED manufactures recommend a flicker frequency above 200 Hz. One last thing that limits \( f_{\text{min}} \) is the background interference. The major energy of an image is on the low frequency range close to DC frequency. The sharp details in the image contribute higher frequency components. Signal could be polluted by the background scene when the background is nonuniform and under strong ambient light. Fig. 3.4 illustrates how background scene interferes the received signal. In this example, a nonuniform wallpaper is used to show the background interference and a 800 Hz signal is transmitted. \( f_{\text{min}} \) should be chosen to avoid conflicting with background scene or background removal operation is required to suppress the interference.

3.2.2 Frequency upper bound of optical tags

Next we explore the upper bound of detectable frequency \( f_{\text{max}} \). According to Nyquist theorem, in order to recover the signal from its sampled version, the sampling frequency \( f_s \) should be at least double of the signal bandwidth. In our case, the measured \( f_s \) is 79 KHz, indicating a maximal 39.5 KHz detectable signal frequency. However, as can be seen from Fig. 3.5, the frequency response of the camera suffers a significant drop after 1 KHz. This is because the effect of exposure time to the received signal has not been taken into consideration yet. To learn about the effect of exposure time, we analyze three different cases shown in Fig. 3.6. We define the contrast as the optical intensity difference between the brightest row and the darkest row and assume the LED power is \( p_{\text{LED}} \) when
Figure 3.4: (a) Received 800 Hz signal with nonuniform background and ambient. (b) The FFT amplitude of pure background scene. (c) The FFT amplitude of modulated signal combined with background scene. (d) The FFT amplitude of received signal after background removal operation.
LED is on. In Fig. 3.6a, the transmitted signal period is longer than the exposure time and the contrast is $p_{LED} \cdot \tau_{exposure}$. In comparison, the contrast is zero in Fig. 3.6b, where the exposure time $\tau_{exposure}$ is exact multiple times of signal period. This is the worst case since the accumulated optical energy of each row is the same and the signal is averaged out. When the signal period is less than $\tau_{exposure}$ (Fig. 3.6c), the contrast is $p_{LED} \cdot \tau_{readout}$ for this case. Since $\tau_{exposure}$ is usually much larger than $\tau_{readout}$, Fig. 3.6a has a much higher contrast than Fig. 3.6c. This explains the frequency response curve in Fig. 3.5. In our measurement, the fastest shutter speed is $1/1000$ s limited by the application. When the transmitted signal exceeds 1 KHz or its frequency is multiple of 1 KHz, the received signal suffers a significant attenuation. Therefore, the upper bound $f_{max}$ is limited by shutter speed of smartphone.

Figure 3.5: Frequency response of camera

(a) signal period $> \tau_{exposure}$    (b) signal period $= \tau_{exposure}$    (c) signal period $< \tau_{exposure}$

Figure 3.6: Effect of exposure time to the received signal.
3.3 Frequency detection

The fundamental frequency detection (or period detection equivalently) can be accomplished in the time domain or in the frequency domain. Algorithms applied in the time domain vary from simple zero-crossing methods to sophisticated auto-correlation based methods. While frequency domain based approaches usually investigate the Fourier spectrum, time domain based approaches can achieve accurate detection for highly periodic signal in low-noise environment. However, these detection algorithms can sometimes fail with noisy signals. In our application, the signal can be highly polluted by the background and severely distorted by the unbalanced illumination, so frequency spectrum analysis is used here.

In order to find the fundamental frequency of the signal $s(t)$, fast Fourier transform (FFT) can be applied to find the peak amplitude in frequency domain. As we know, the accuracy depends on the length of the signal used to calculate the FFT [4]. Because FFT corresponds to samples of discrete-time Fourier transform (DTFT) at frequency $k/N$, where $N$ is the number of points used to calculate FFT and $k \in \{0, 1, ..., N - 1\}$, the frequency accuracy is determined by length $N$. An sampling frequency $f_s$ (Hz) will lead to an accuracy of $f_s/N$ Hz. And when the sampling frequency $f_s$ is not exact multiple of the signal’s fundamental frequency, the FFT will miss the peak point of DTFT. The DTFT should be used to find the peak in frequency domain instead of FFT.

As all the possible frequencies $f_i$ of the transmitted signals are pre-known at the receiver, determining the LED optical tag is actually a frequency detection problem instead of frequency estimation. A single DTFT value at frequency $f_i$ can be calculated to determine
the LED ID corresponding to $f_i$. The reliability of detection and the frequency resolution when multiple frequency occur simultaneously are our main concerns here.

### 3.3.1 Reliability of frequency detection

First, we discuss the reliability of frequency detection under noise environment. Noise brings two major impacts to the frequency measurement: 1) it smears the peak of DTFT spectrum at frequency $f_i$, making it difficult to identify the presence of signal; 2) it causes the peak point deviating from true frequency, degrading the measurement accuracy. Understanding the noise effect to the frequency measurement is important to the frequency detection. Since the transmitted square waveform $s(t)$ is a periodic signal with fundamental frequency $f_i$, it can be expressed using discrete Fourier series:

$$s[n] = \sum_{k=-\infty}^{\infty} \alpha_k e^{j2\pi nk f_i / f_s}$$

Considering that the most energy of the is on the fundamental frequency $f_i$, to simplify analysis, we only examine the exponential component on fundamental frequency $f_i$. So we define the signal model as:

$$y[n] = \alpha e^{j\theta_i n} + u[n]$$  \hspace{1cm} (3.1)

where $\theta_i = 2\pi f_i / f_s$ and $u[n]$ is assumed to be a white noise with zero mean and variance $\sigma^2$ for simplification. To reduce the side lobe effect, a weighting window is usually applied before DTFT. Multiplying a window $w[n]$ to Eq. 3.1 we have

$$x[n] = y[n]w[n] = \alpha e^{j\theta_i n}w[n] + u[n]w[n]$$  \hspace{1cm} (3.2)
The DTFT of windowed signal is given by

\[
X(\theta) = \alpha \sum_{n=0}^{N-1} w[n]e^{j(\theta_i - \theta)n} + \sum_{n=0}^{N-1} w[n]u[n]e^{-j\theta n}
\] (3.3)

The DTFT value at angular frequency \(\theta_i\), which corresponding to frequency \(f_i\) can be calculated as

\[
X(\theta_i) = \alpha \sum_{n=0}^{N-1} w[n] + \sum_{n=0}^{N-1} w[n]u[n]e^{-j\theta_i n}
\] (3.4)

The mean of the square magnitude of the DTFT value at \(f_i\) is given by

\[
E(|X(\theta)|^2) = (\alpha \sum_{n=0}^{N-1} w[n])^2 + \sigma^2 \sum_{n=0}^{N-1} w^2[n]
\] (3.5)

The derivation of the results beyonds the scope of this paper, only the final formula is given here. Please refer to book [39] for more details about the derivation. We define the signal-to-noise ratio (SNR) as windowed signal power at frequency \(f_i\) to the noise power at frequency \(f_i\):

\[
SNR = \frac{(\alpha \sum_{n=0}^{N-1} w[n])^2}{\sum_{n=0}^{N-1} w[n]u[n]e^{-j\theta_i n}}
\] (3.6)

From Eq. 3.6 we can see that the SNR depends on both the window types and number of points used to calculate DTFT. In practice, a 14 dB SNR (25 in absolute value [43]) usually assures a reliable detection with a high probability of signal presence.
3.4 resolution of frequency detection

Frequency resolution is another crucial factor. It decides the minimum frequency spacing between two neighboring frequencies, which is related to the system capacity. The frequency resolution is defined as the ability to distinguish the two different frequency close to each other when multiple LED IDs occur at the same time and captured by one camera frame.

Due to the limited number of sampling points \( N \) to calculate the Fourier spectrum, unwanted lobes and ripples appear when using DTFT to analyze the spectrum. The exist of lobes in spectrum will smear the peak points and affect the frequency detection reliability. Assuming we have two neighboring frequencies \( f_i \) and \( f_{i+1} \) so close to each other that it is difficult to separate them in spectrum because of lobes, we are not confident to decide whether the frequency \( f_{i+1} \) occurs or not without additional information because the DTFT peak at \( f_{i+1} \) could result from the lobes of \( f_i \). This is especially true when the amplitudes of these two frequencies differs a lot. Since the lobes are caused by the limited length of window applied to the signal, the frequency resolution also depends on the window length and window type. We define the frequency resolution \( f_{res} \) as the main lobe width of window. With this separation, the lobe's magnitude attenuates deep enough to distinguish two different frequencies. A few windows’ main lobe width and side lobe attenuation are given in table 3.1. The interval between neighbouring \( f_i \) should be at least the main lobe width of the window in order to successfully separate them from the received image.
Table 3.1: Window Characteristics

<table>
<thead>
<tr>
<th>Window Type</th>
<th>$-3$ dB Main Lobe Width (bins)</th>
<th>Main Lobe Width (bins)</th>
<th>Side Lobe Attenuation(dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rectangular</td>
<td>0.89</td>
<td>$2/N$</td>
<td>$-13$</td>
</tr>
<tr>
<td>Hann</td>
<td>1.44</td>
<td>$4/N$</td>
<td>$-31$</td>
</tr>
<tr>
<td>Hamming</td>
<td>1.30</td>
<td>$4/N$</td>
<td>$-41$</td>
</tr>
<tr>
<td>Blackman</td>
<td>1.64</td>
<td>$6/N$</td>
<td>$-61.5$</td>
</tr>
</tbody>
</table>

3.5 Background subtraction

Most of current work only examine the cases with close range to modulated light and uniform reflecting surface under no ambient light interference, which is usually not the case in practice. When the background is not uniform, background removal is a necessary procedure. This is usually the case where the modulated illumination is not the only light source and the ambient light could not be ignored. Even though most energy of an image is in low frequency range, the sharp detail in the image could contribute high frequency components to frequency range of interest. A simple way to deal with this is defocusing, which acts as a low-pass filter that compresses the the high frequency component of the image to some extent while retaining the LED beacon ID information. However, the defocus function depends on the support of smart phone operating system application program interface and could not effectively remove the interference from background [41]. Bytelight [11] averages a sequence of video frames to get a pure background image without modulated illumination. As long as the modulation frequency is not an exact multiple of the video frame rate, the position of bands will shift in different frames. An average process will cancel out the effects of bright and dark bands, leaving a background with little rolling shutter effect. Subtracting the background from an image might increase the contrast of
rolling shutter effect, but its performance heavily depends on the illumination condition of different images in order to recover the signal. Besides, the requirement of image alignment to each frame of the video sequence brings in a significant process delay and computation complexity. Furthermore, the following discussion shows that direct subtracting will not promise good separation of modulated illumination and background. In this letter, we propose a background removal algorithm using one additional image to recover the modulated information from background scene, which can greatly compress the background and ambient light interference. Linear filter performs poorly in this case because the image with rolling shutter effect is the multiplication of modulated signal and the reflectance of the scene. Therefore we propose background division rather than background subtraction to remove the background interference.

Since the major energy of background scene is in DC frequency range, one may think using a high pass filter to extract polluted frequency components from background. However, linear filtering performs poorly in this application since the reflectance and modulated illumination were original combined by multiplication, not addition. Linear filtering cannot correctly separate the signals with a nonlinear operation [43]. To understand this, think about how cameras form an image. The image captured by the camera is formed from the light bouncing off of an object that does not emit light itself. This means the acquired image is the result of the objects reflectance multiplied by the light source. The intensity of incident light reaching on the objects may be different, but the percentage of the incident light reflected from the objects is the same. The goal of the background remove is to recover the modulated illumination, despite of background object reflectance.
Figure 3.7: (a) nonuniform reflectance of the background; (b) modulated illumination; (c) captured image with rolling shutter effect; (d) average of a sequence of frames of images with rolling shutter effect; (e) directly subtract Fig. 3.7d from Fig. 3.7c; (f) divide Fig. 3.7c by Fig. 3.7a.
Fig. 3.7 shows how this works. Fig. 3.7a represents the nonuniform reflectance of the background, which could be a wallpaper in real world. Fig. 3.7b illustrates an example of modulated illumination reaching on Fig. 3.7a. Fig. 3.7c is the camera acquired image, which is equal to the background reflectance, multiplied by the modulated illumination. Fig. 3.7d is the averaged image of 40 frames with different delay to cancel out the modulated illumination. Fig. 3.7e is the result after directly subtracting Fig. 3.7d from the viewed image Fig. 3.7c. It can be seen from this figure that most sharp details of background scene still remain. Fig. 3.7f is obtained from a nonlinear process, dividing Fig. 3.7c by Fig. 3.7a. According to the result, we can see that Fig. 3.7f is a better approximation of Fig. 3.7b.

Back to the rolling shutter based VLC model after the discussion of the example in Fig. 3.7, our goal is to recover the modulated illumination from the viewed image, given an approximation of background reflectance. Suppose $s(x, y)$ ($x \in [1, ..., M]$ and $y \in [1, ..., N]$), where $M$ and $N$ is the dimension of the acquired image) is a periodic signal with frequency $f$ used as the LED ID beacon. $s(x, y)$ is usually a unipolar square waveform which can be generated by a low-cost and simple structure driver. Since the optical intensity can not be negative value, $s(x, y)$ is a unipolar signal with a DC offset. And the ambient light is represented by a constant $\rho$. In fact, $\rho$ is not a constant value and usually depends on the positions of specific pixel due to the variation of ambient light source power distribution as well as the distance between objects and ambient light source. However, the small fluctuation of $\rho$ can be negligible and a constant value is a good approximation in our discussion of the model. So acquired image $p(x, y)$ containing the rolling shutter effect is equal to the reflectance of background scene $b(x, y)$, multiplied by the sum of the signal
$s(x,y)$ and the DC offset $\rho$.

$$p(x,y) = b(x,y) \times [\rho + s(x,y)] \quad (3.7)$$

The noise is not our focus here and is omitted. Here we assume the entire image is formed from reflecting light and contains no active light source. This model is still valid when including active light source by allowing reflectance larger than 1.

Next step is to find a good approximation of the reflectance of the background scene. As we know the observation of rolling shutter is determined by the exposure time. By setting the exposure time to auto mode, a background scene without rolling shutter effect can be easily acquired. The background scene $p'(x,y)$ can be approximated as the multiplication of background reflectance $b(x,y)$ and a constant light intensity $\rho'$:

$$p'(x,y) = b(x,y) \times \rho' \quad (3.8)$$

Generally, $\rho'$ is larger than $\rho$ since it accumulates more energy due to longer exposure time. Because of the movement of mobile user, image alignment might need to be applied to the background scene. Image processing, such as motion compensation and feature matching, could be used to align these two images. After image alignment, the non-overlapped part of these two images could be discarded while retaining the overlapped part. In our proposal, only one extra image is needed to get the background scene and only one time image alignment is required. To separate the modulated illumination, the acquired image with rolling shutter effect should be divided by the reflectance of background scene. Divide
Eq. 3.7 by Eq. 3.8, we get
\[ \frac{p(x, y)}{p'(x, y)} = \frac{\rho}{\rho'} + \frac{s(x, y)}{\rho'}. \]  \hspace{1cm} (3.9)

The term \( \frac{\rho}{\rho'} \) represents a DC offset while \( \frac{s(x, y)}{\rho'} \) is the signal with its amplitude rescaled to \( 1/\rho' \). This nonlinear process can be converted into a linear one equivalently by taking the logarithm of both images. Subtraction can be applied after taking logarithm to isolate the illumination.

Since the rolling shutter effect occurs in a column pattern, the two dimensional \( M \times N \) image can be converted to a size \( N \) array to reduce the computation while keeping the signal information, before we take further process. The mean of each column is calculated to form a one-dimension array which contains the modulated frequency information. A polynomial curve fitting gives a good approximation of the small variation of ambient light trend [7, 5]. Subtracting the polynomial fitted value from Eq. 3.9 will pull \( s(x, y) \) back to same reference level and remove major DC component. After this, Fourier spectrum analysis can be used to detect the appearance of the LED ID beacon.

The procedure can be summarized in algorithm 1

\begin{algorithm}
\caption{Background removal}
1: procedure
2: \hspace{1cm} Set different exposure time to take a signal image and background image;
3: \hspace{1cm} Convert both images to gray scale and align two images using image processing techniques;
4: \hspace{1cm} Divide signal image by the background image and convert it to a one-dimension array;
5: \hspace{1cm} Subtract the polynomial fitted value to eliminate major DC component;
6: \hspace{1cm} Apply the Fourier spectrum analysis to detect the appearance of the LED ID beacon
7: end procedure
\end{algorithm}
Experiments are conducted to validate this theory. Fig. 3.8a is the background captured by setting the exposure time to normal. The background contains a stripe pattern with fundamental frequency around 290 Hz to interfere the modulated illumination on purpose. Fig. 3.8b is the signal image with rolling shutter effect. It can be seen from Fig. 3.8b, the sharp details of background retains due to the strong ambient light. The modulated illumination pattern is not easily recognized from the background scene.

Fig. 3.9 shows the Fourier spectrum of different images. All the amplitude of fast Fourier transform (FFT) has been normalized. The amplitude of FFT of pure background scene is displayed in Fig. 3.9a. We can see two separate peaks located at 290 Hz, which is the fundamental frequency, and 870 Hz, which is its third order harmonic frequency. Fig. 3.9b illustrates the FFT spectrum of signal image with rolling shutter effect. The frequency of modulated illumination is 800 Hz. We can see that the signal is submerged by the background scene and is not detectable. Linear filter will not work since the background
Figure 3.9: Fourier spectrum analysis

(a) Fourier spectrum of background image

(b) Fourier spectrum of signal image with rolling shutter effect

(c) Fourier spectrum of signal image after background remove
is mixed with modulated illumination in the same frequency range of interest. Fig. 3.9c shows the frequency spectrum of signal image after background remove. We can clearly see a peak at around 800 Hz frequency and the background interference is significantly suppressed. In this case, the LED ID beacon become detectable.
Chapter 4

Multi-view Geometry Positioning

with Extended Kalman Filter

In this chapter, we introduce proposed VLC based positioning algorithm using
multi-view geometry with built-in inertial sensors. We first give an overview of the sensors
built inside smartphones. Then the multi-view geometry using single camera is investigated
with additional information provided by those built-in sensors. Finally, an extended Kalman
filter is applied to maximize the probability of correct estimation of system states.

4.1 Smartphone Sensor Overview

Nowadays, smartphone manufactures have integrated all kinds of different sensors
into one cellphone to make their smartphone more powerful and smarter. These built-in
sensors are capable of measuring motion, orientation and various environmental conditions
with high precision and accuracy. It enables the smartphone to provide complicated appli-
4.1.1 Motion sensors

Motion sensors provide raw data that measures acceleration forces and rotational forces along its three axes. Motion sensors include accelerometers and gyroscopes. However, some manufactures may also provide gravity sensors and rotational vector sensors. The later two sensors can be physical hardware sensors or virtual software-based sensors derived from one or more physical sensors. These measured data are with respect to the sensor’s own coordinate system as shown in Fig. 4.1.

Accelerometer: the accelerometer can sense the change of cellphone orientation by
measuring the acceleration force \( (m/s^2) \) applied to the device along its three physical axes, including the contribution of gravity. Therefore, the stationary device gives a measurement of \( g = 9.81 m/s^2 \) while it reads \( g = 0 m/s^2 \) when the device is in free fall. In order to get a correct acceleration readings, the gravity force must be compensated where a high-pass is usually applied. Generally, accelerometer can monitor the device motion change while maintains a 10 times less power consumption compared to other motion sensors.

Gyroscope: while the accelerometer measures the linear acceleration of movement, the gyroscope measures the rate of the angular velocity \( (rad/s) \) around its \( x, y \) and \( z \) axis. It can detect the motion of roll, pitch and yaw. In practice, the relative change of the angles around three axes are calculated by integrating the gyroscope over time.

Gravity sensor: it measures the acceleration caused by the gravity along its three axes.

Rotational vector sensor: this kind of sensor evaluates the rotation matrix from sensor coordinates system to global coordinates system.

The raw data provided by sensors contains noise and bias that introduce errors. In practice, the bias and noise can be compensated by integrating data from multiple sensors.

4.1.2 Position sensors

Magnetometer and orientation sensors are position sensor that monitor the device position.

Magnetic field sensor: magnetometer provides raw measurement of magnetometer data along three physical axes and a computed compass bearing. One of the main function of magnetometer is to give device pose with respect to the north-south pole of earth by use
of magnetism. Magnetic field sensor uses coordinates systems relative to earth magnetic field (Fig. 4.2).

Proximity sensor: proximity sensor is usually comprised of an IR transceiver, near the earpiece speaker. It measures the proximity of object relative to the device screen, to turn off the screen when the handset is being held up.

4.1.3 Environmental sensors

Sensors of this category provide information about the user environment like ambient illumination level, temperature, air pressure and humidity. Environmental sensors include photometer, thermometer and barometer.

Photometer: photometer measures the ambient illumination (in lx) to adjust screen display brightness. It measures light intensity over different spectrum and combine the result to eliminate the source dependency (eg. lamp that emits light from certain spectrum). It has low light sensitivity, dynamic range and adjustable gain.
Thermometer: as the name indicates, thermometer measures the ambient temperature (in degree Celsius).

Barometer: barometer measures atmospheric pressure and is usually used in forecasting the weather and determining altitude. The altitude information can be used to improve GPS accuracy.

Besides these sensors, some smartphones also equip dedicated sensors for healthy and fitness applications. For example, the pedometer counts the number of steps of the mobile user. The heart rate measure the pulsations of blood vessels inside the finger to monitor heart rate.

4.2 Camera Model

In order to transform a physical global 3D coordinates to a 2D image plane, a camera model needs to be developed to analyze the imaging geometry.

4.2.1 Pinhole camera model

A simple but useful camera model is the pinhole camera model [3]. A pinhole model assumes that the aperture of the camera is a tiny hole in the center of the camera optical axis, which blocks all incoming light except those travelling through the tiny hole. As is shown in Fig. 4.3, \( f \) is the focal length, \( X \) is the object length and \( Z \) is the distance to the pinhole. The position of projected point \( x \) can be solved using similar triangles:

\[
x = -f \frac{X}{Z}
\]  

(4.1)
Figure 4.3: Pinhole camera model: a pinhole aperture projects only those rays intersected with optical axis to image plane.

In our idealized model, assumption that the image plane is always in focus is made. So the position of projected point is only depend on one parameter: the focal length of camera. For math convenience, the pinhole plane and image plane are rearranged in Fig. 4.4 to form an equivalent model. The main difference is that the imaged object has the same orientation as the physical object and the negative sign is eliminated. The camera coordinates system is also illustrated in Fig. 4.4. The pinhole, reinterpreted as the center of projection, is the origin of the camera coordinates system and the $z$ axis is along the optical axis. A physical point $P (X, Y, Z)$ with respect to the camera coordinates now is projected to a point $p (x, y, f)$ in the image plane. And the $x$ and $y$ can be represented as:

\[
\begin{align*}
  x &= f \frac{X}{Z} \\
  y &= f \frac{Y}{Z}
\end{align*}
\]  

(4.2)
Figure 4.4: Rearrangement of image plane and aperture plane. The point $P$ is projected to the image plane $p$, the negative sign is eliminated by this rearrangement.

Usually, the image has its own 2D coordinates system $(u, v)$, and the origin of the coordinates is not on the optical axis (e.g., left topmost pixel is the origin). This will introduce two new parameters $c_x$ and $c_y$, representing the displacement (in pixel) from the optical axis to origin of 2D image coordinates. The position $(u, v)$ of the projected point can be calculated in accordance with the following equations:

$$
\begin{align*}
    u &= f_x \left( \frac{X}{Z} \right) + c_x \\
    v &= f_y \left( \frac{Y}{Z} \right) + c_y
\end{align*}
$$

(4.3)

The $f_x$ and $f_y$ are the focal length in $x$ axis and $y$ axis respectively with unit in pixel. We use two different focal length because in general the pixel size of the image sensor is not square but rectangular. This focal length with units in pixel is the product of physical lens focal length and the dimension of individual pixel of image sensor. $f_x$, $f_y$, $c_x$ and $c_y$ are called camera intrinsic parameters.
4.2.2 Lens distortions

However, the pinhole model cannot be used for imaging in practice since it will not collect enough light for fast exposure. This is the reason why camera lens is used to gather more light reaching on the image sensor. The using of camera lens not only makes the imaging geometry more complicated but also brings distortions to the captured image. Radial distortions and tangential distortions are two distortions occurring in imaging systems that mainly affect the imaging performance.

The radial distortion results from the distortion of lens sharp, and is commonly recognized as fish-eye effects. Rays at the edge of an image are bent more than those close to the center of the image. The distortion is negligible at the optical axis and increases when moving toward the edge. For highly distorted fish-eye cameras lens, a third order of Taylor series expansion is good enough describe the radial distortion [3]:

\[
x_{corrected} = x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\
y_{corrected} = y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6)
\] (4.4)

where \( r \) is the radius to the optical axis.

The tangential distortion is caused by the misalignment of the lens and image sensor during the assembling process of camera. This distortion can be characterized by two parameters \( p_1 \) and \( p_2 \):

\[
x_{corrected} = x + [2p_1 y + p_2(r^2 + 2x^2)] \\
y_{corrected} = y + [p_1(r^2 + 2y^2) + 2p_2 x]
\] (4.5)
Figure 4.5: Radial distortion: rays at the edge of an image are bent more than those close to the center of the image, known as barrel distortion or fish eye distortion.

Figure 4.6: Tangential distortion caused by the misalignment of the lens and image sensor during the assembling process of camera.

Thus there are five distortion related parameters in total need to be calculated before we apply imaging geometry analysis.

4.2.3 Camera pose

In general, the location of a physical object is expressed with coordinates with respect to the global coordinates system as shown in Fig. 4.7. To convert the 3D coordinates into 2D coordinates, we need to transform the global coordinates system to the camera
coordinates system first. The transformation can be operated by multiplying a $3 \times 3$ rotation matrix $C^G R$ to the point $G^P$ with respect to the global coordinates system:

$$C^P = C^G R^G P$$

Where $C^P$ is the corresponding point in the camera coordinates system. The rotation matrix $C^G R$ can be characterized by three angles $\psi$, $\varphi$ and $\theta$ that representing rotation around $x, y$ and $z$ axis respectively:

$$C^G R = R_x(\psi) R_y(\varphi) R_z(\theta) \quad (4.6)$$

where

$$R_x(\psi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & \sin \psi \\ 0 & -\sin \psi & \cos \psi \end{bmatrix}$$

$$R_y(\varphi) = \begin{bmatrix} \cos \varphi & 0 & -\sin \varphi \\ 0 & 1 & 0 \\ \sin \varphi & 0 & \cos \varphi \end{bmatrix}$$

$$R_z(\theta) = \begin{bmatrix} \cos \theta & \sin \theta & 0 \\ -\sin \theta & \cos \theta & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

One thing need to be noted is that the rotation around three axes must be done in a
Figure 4.7: Converting point in global coordinates system to camera coordinates system: the point $P$ in the global frame is projected to the image plane as point $p$; the point $p$ is related to point $P$ by applying a rotation matrix and a translation vector in a certain sequence. Otherwise, different rotation sequence will cause different corresponding rotation matrix. The transpose of matrix $C_R G_T$ is the inverse operation which convert a point in camera coordinates system to a point with respect to the global coordinates.

Besides the three angles representing the rotation around three axes, a translation vector $T = (X_T, Y_T, Z_T)$ is also needed to donate the shift from the origin of global coordinates to the origin of camera coordinates as shown in Fig. The offset $T$ can be simply calculated in accordance to $T = G_{\text{origin}} - C_{\text{origin}}$. The three rotation parameters and three offset parameters are called extrinsic parameters. A point $C_P$ in camera coordinates, which corresponding to a point $G_P$, can be represented as:

$$C_P = C_R G_T (G_P - T)$$  \hspace{1cm} (4.7)

Brings Eq. 4.7 into Eq. 4.3, then we can convert a 3D point in global coordinates...
into a 2D point on the image plane.

As discussed above, the camera pose introduces additional six parameters (three rotation angles and a shift vector). With four camera intrinsic parameters \((f_x, f_y, c_x\) and \(c_y)\) and five distortion coefficients (three radial coefficients and tangential coefficients), there will be fifteen parameters in total to be resolved in order to transform a 3D point into a 2D point. Good thing is, the four camera intrinsic parameters and five distortion coefficients can be solved through a camera calibration process [55]. However, the six extrinsic parameters changes between every frame. It is noted that the translation vector \(T\) is what we need to solve for positioning purpose.

4.3 LED recognition on the image plane

Generally, the LED will be the brightest pixels since it directly cast light to these pixels on the image sensor. In order to find the LED candidates, the intensity can be threshold to convert the image into a binary image:

\[
I(u, v) = \begin{cases} 
1, & \text{if } I(u, v) > \tau \\
0, & \text{otherwise}
\end{cases}
\]

Pixel intensity greater than threshold will be set to 1 and others will be set to 0. Existing image processing techniques like contour detection and blob detection can be used to extract the LED candidates from the binary image. After that, the detected contours can be filtered according to its area, radius and shape. However, false detection could still exists because of the appearance of unmodulated light source. To further refine the result, FFT will be
applied to the remaining contours to detect the frequency component using rolling shutter effects. The projection of the modulated LED on the image plane should be extracted in this way. The geometry centroid of the contour is used as the position of LED on image plane. Compared to camera frame based visible communication, the advantage of using rolling shutter is that there is no chance to capture an image with LED in off state.

### 4.4 Get orientation from accelerometer and magnetometer

In the following sections, we assume that the camera is well calibrated such that the four camera intrinsic parameters \((f_x, f_y, c_x, c_y)\) as well as the five distortion coefficients \((k_1, k_2, k_3, p_1, p_2)\) are already known. For a one camera single view geometry, at least three LEDs with position known are needed to calculate the mobile device position if the orientation of the mobile device is not known. However, if we use smartphone as our receiver, then those built-in sensors can provide more information to eliminate some unknown variables hence reducing the numbers of required LEDs. We will find that in the following discussion, the orientation of the mobile device can be sensed using an accelerometer and magnetometer.

The output of the sampled accelerometer measurement is the following continuous-time signal:

\[
\mathbf{a}_m(t) = \mathbf{I}_{\mathcal{G}} \mathbf{R} (\mathbf{G}_a \mathbf{a}_f(t) - \mathbf{G}_g \mathbf{g}) + \mathbf{b}_a(t) + \mathbf{n}_a(t)
\]  

(4.8)

where \(\mathbf{G}_a \mathbf{a}_f(t)\) and \(\mathbf{G}_g \mathbf{g}\) denote the linear acceleration and gravity in global frame, \(\mathbf{b}_a(t)\) and \(\mathbf{n}_a(t)\) represent the bias and white Gaussian noise process. Eq. 4.8 implys that accelerometer measuring forces, including both body and gravity forces, on its own three axes expressed
in the local frame. It is further assumed that:

- The mobile device is in static hence there is no linear acceleration ($^G a_I(t) = 0$).
- The bias is compensated so that $b_a(t) = 0$
- The noise is negligible ($n_a(t) \approx 0$)

With these assumptions, Eq. 4.8 becomes:

$$a_m(t) = -^I_G R_g^G g$$

The rotation matrix $^I_G R$ is the same in Eq. 4.6, then Eq. 4.9 becomes:

$$a_m(t) = -
\begin{bmatrix}
    \cos \varphi \cos \theta & \cos \varphi \sin \theta & -\sin \varphi \\
    \cos \theta \sin \varphi \sin \psi - \cos \psi \sin \theta & \cos \psi \cos \theta + \sin \varphi \sin \psi \sin \theta & \cos \varphi \sin \psi \\
    \cos \psi \cos \theta \sin \varphi + \sin \psi \sin \theta & \cos \psi \sin \varphi \sin \theta - \cos \theta \sin \psi & \cos \varphi \cos \psi
\end{bmatrix}
\begin{bmatrix}
    0 \\
    0 \\
    1
\end{bmatrix}g
$$

It can be seen that Eq. 4.10 only depends on two rotational angles $\varphi$ and $\psi$ thus can be solved. The remaining unknown rotational angle $\theta$ around $z$-axis is nothing but the rotation from magnetic north, so it can be determined using a magnetometer sensor. More details of the operation of calculate device orientation from accelerometer and magnetometer
sensor can be found in [38]. Here we assume that the device is in static so that linear acceleration $G_a(t) = 0$. In fact, $G_a(t)$ can be filtered out using a low-pass filter while remaining the gravity components as we discussed in previous section.

4.5 Positioning using multi-view geometry

According to the previous analysis of camera model, in order to calculate the mobile device location, the six extrinsic parameters must be solved after the camera is well calibrated. At least three LEDs with pre-known position are required since one projection of LED can derive two equations [52]. If the accelerometer sensor and the magnetometer can be used to estimate the orientation of camera respect to the global coordinates system, then the three rotation angles $\psi$, $\varphi$ and $\theta$ can be solved. In this case, there are three remaining parameters in translation vector $T = (X_T, Y_T, Z_T)$ which directly related to the mobile device position need to be estimate. The transformation from 3D world to 2D image lose the depth information, so these three parameters still need two projection of LEDs to solve the translation vector. However, in practice the image sensor can usually have only one LED in its view during most of the time due to the narrow FOV of front camera and short distance to the ceiling. A naive solution is to eliminate one unknown variable $Z_T$ by assuming the height of mobile device is a constant value. In this case, one LED is enough to calculate the position of mobile device. Another method is that, instead of using multiple LEDs, two cameras with relative orientation and position known can be applied to determine the position using vision triangulation [51]. However, there is only one front facing camera on most smartphone except those customized.
Our objective is to design a VLC positioning algorithm with following features:

1) The positioning algorithm can work with only one detected LED and can seamlessly switch between one LED mode and multi-LED mode;
2) The positioning algorithm can still estimate device location when the VLC link is temporarily blocked or the LED temporarily moves out of the FOV of the camera;
3) No additional accessory should be added to the smartphone except those sensors built inside the phone;
4) The position of the mobile device is calculated at the receiver end. Hence, there is no privacy concern;
5) The algorithm should not depend on the knowledge of transmitter like shape, radiation and transmitted power;
6) The orientation and height of the mobile device should not be assumed to be fixed.

Inspired by the multi-camera vision triangulation algorithm, we propose a VLC positioning algorithm using multi-view geometry. The idea is that once the offset between two camera frame can be determined, then the position of mobile device can be also determined. In our algorithm, accelerometer and gyroscope sensors are adopted to monitor the translation and pose changes between two camera frames.

First, we need to initialize the the pose and position of device at the first time instant using the camera measurement of image frame coordinates of one LED, as long as the LED is in the FOV of the camera. To accomplish the initialization process, the smartphone is required to be held steady. The purpose to do this is: 1) manually set the initial linear speed as 0; 2) get more accurate pose estimation using method introduced in Sec. 4.4. The transformation of LED in the global frame to the camera coordinates are
given by:

\[
\begin{bmatrix}
    C_{x_k} \\
    C_{y_k} \\
    C_{z_k}
\end{bmatrix} = \mathcal{C}_I^G R_k (\mathbb{G} p_L - \mathbb{G} p_{I_k}) + \mathbb{C}_p \tag{4.11}
\]

with \([C_{x_k} \ C_{y_k} \ C_{z_k}]^T\) being the LED position with respect to the camera at time \(k\), \(\mathcal{C}_I\) and \(\mathcal{C}_G\) is the rotation matrix that transform point in IMU frame to camera frame and rotation matrix from global frame to IMU frame at time instant \(k\), respectively. And \(\mathbb{G} p_L\) and \(\mathbb{G} p_{I_k}\) is LED and IMU position at time \(k\) in the world coordinates, respectively. \(\mathbb{C}_p\) represents the translation from the IMU to camera in camera frame since the center of camera and center of IMU are usually not overlapped. Bring Eq. 4.11 into Eq. 4.3 we can get the projection of LED onto the image plane once the camera is calibrated. But for math convenience, we will use Eq. 4.11 as our camera measurement instead of Eq. 4.3.

When camera acquires next frame of measurement of LED position at time \(k+1\), there will be a new measurement equation:

\[
\begin{bmatrix}
    C_{x_{k+1}} \\
    C_{y_{k+1}} \\
    C_{z_{k+1}}
\end{bmatrix} = \mathcal{C}_I^G R_{k+1} (\mathbb{G} p_L - \mathbb{G} p_{I_k} - \mathbb{G} V) + \mathbb{C}_p \tag{4.12}
\]

where the vector \(\mathbb{G} V\) represents the translation from the position at time \(k\) to the position at time \(k+1\). \(\mathbb{G} V\) can be calculated using data provided by accelerometer and gyroscope. Eq. 4.11 and Eq. 4.12 give two different views of the same LED, then the vision triangulation can be used to compute the mobile device position. The calculation of translation \(\mathbb{G} V\) will
be discussed in the following section.

4.5.1 Kinematic equations

The measurements of gyroscope and accelerometer can be modeled in a continuous-time form:

\[ w_m(t) = I w(t) + b_g(t) + n_r(t) \] (4.13)

\[ a_m(t) = I_G R (G a_I(t) - G g) + b_a(t) + n_a(t) \] (4.14)

Eq. 4.14 is the same as Eq. 4.8, we rewrite here for reference. Eq. 4.13 shows the gyroscope measures the rotational velocity along its three axes, with \( I w(t) \) denote the angular rate in its own frame, \( b_g(t) \) is the gyroscope bias and \( n_r(t) \) is the Gaussian noise. We assume that what the IMU data provides is the sampled version of \( w_m(t) \) and \( a_m(t) \) at discrete time step \( t_k \) (\( k = 0, 1, 2, \ldots \)). In what follows, we will show how \( w_m(t) \) and \( a_m(t) \) can be used to monitor the global pose, velocity and position from the transformation of local motion information.

Gyroscope propagation

There are several different forms to represent the orientation change of the IMU frame from time \( t_k \) to time \( t_{k+1} \), like Euler angles, rotation matrix and quaternion. Here, we follow the convention in [45], using a unit quaternion \( I_{k+1} q \) to represent the relative rotation of the IMU between \( t_k \) and \( t_{k+1} \). To estimate \( I_{k+1} q \) using \( w_m(t) \), we first obtain...
the rotational velocity between \([t_k, t_{k+1}]\) after the bias compensation:

\[
\mathbf{w}(t) = \mathbf{w}_m(t) - \mathbf{b}_g(t)
\]  

(4.15)

and then the quaternion derivative can be expressed as follow:

\[
\begin{bmatrix}
\frac{I_k^T}{I_k} \mathbf{q} \\
\end{bmatrix} = \frac{1}{2} \begin{bmatrix} -[\mathbf{w}(t) \times] & \mathbf{w}(t) \\ -\mathbf{w}(t)^T & 0 \end{bmatrix} \begin{bmatrix}
\frac{I_k^T}{I_k} \mathbf{q} \\
\end{bmatrix}
\]  

(4.16)

with the initial condition \(\frac{I_k^T}{I_k} \mathbf{q} = \begin{bmatrix} 0 & 0 & 0 & 1 \end{bmatrix}^T\). \([\mathbf{w}(t) \times]\) is the skew-symmetric operator, defined as:

\[
[\mathbf{w}(t) \times] = \begin{bmatrix}
0 & -w_3(t) & w_2(t) \\
w_3(t) & 0 & -w_1(t) \\
-w_2(t) & w_1(t) & 0 \\
\end{bmatrix}
\]  

(4.17)

The relative orientation \(\frac{I_{k+1}^T}{I_k} \mathbf{q}\) of the IMU frame then can be computed by integrating the differential equation in Eq. 4.15. Here it is assumed that \(\mathbf{w}_m(t)\) is a continuous-time signal between time interval \([t_k, t_{k+1}]\) thus all values of \(\mathbf{w}_m(t)\) for every \(t \in [t_k, t_{k+1}]\) are available. However, in practice this assumption is not valid since gyroscope only provides discrete-time measurements of \(\mathbf{w}_m(t_k)\) and \(\mathbf{w}_m(t_{k+1})\). Additional assumptions of how the \(\mathbf{w}_m(t)\) evolves from \(t_k\) to \(t_{k+1}\) need to be employed in order to calculate the integration in Eq. 4.15. Here \(\mathbf{w}_m(t)\) is assumed to change linearly between two samples at \(t_k\) and \(t_{k+1}\). This assumption brings in unavoidable approximation errors. These errors could be negligible if the sampling rate is fast enough. With this assumption, the numerical integration is computed using
fourth-order Runge-Kutta to obtain $I_{k+1}^q$.

To calculate $I_{k+1}^q$, we divide $[t_k, t_{k+1}]$ into four time slots $[t_k, t_k + \frac{1}{4} \Delta t]$, $[t_k + \frac{1}{4} \Delta t, t_k + \frac{1}{2} \Delta t]$, $[t_k + \frac{1}{2} \Delta t, t_k + \frac{3}{4} \Delta t]$ and $[t_k + \frac{3}{4} \Delta t, t_{k+1}]$, where $\Delta t = t_{k+1} - t_k$, and apply fourth-order Runge-Kutta to each time slot. With the assumption that $w_m(t)$ changes linearly, we have that:

$$w(t_k + \frac{1}{4} \Delta t) = \frac{3}{4} w(t_k) + \frac{1}{4} w(t_{k+1})$$

$$w(t_k + \frac{1}{2} \Delta t) = \frac{1}{2} w(t_k) + \frac{1}{2} w(t_{k+1})$$

$$w(t_k + \frac{3}{4} \Delta t) = \frac{1}{4} w(t_k) + \frac{3}{4} w(t_{k+1})$$

Here we give an example of how to calculate the integration of first time slot using fourth-order Runge-Kutta with initial $I_k^q = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$.

$$k_1 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k) \times \rfloor & w(t_k) \\ -w(t_k)^T & 0 \end{bmatrix} I_k^q$$

$$w(t_k + \frac{1}{8} \Delta t) = \frac{1}{2} w(t_k) + \frac{1}{2} w(t_k + \frac{1}{4} \Delta t)$$

$$k_2 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k + \frac{1}{8} \Delta t) \times \rfloor & w(t_k + \frac{1}{8} \Delta t) \\ -w(t_k + \frac{1}{8} \Delta t)^T & 0 \end{bmatrix} (I_k^q + \frac{1}{8} \Delta t k_1)$$

$$k_3 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k + \frac{1}{8} \Delta t) \times \rfloor & w(t_k + \frac{1}{8} \Delta t) \\ -w(t_k + \frac{1}{8} \Delta t)^T & 0 \end{bmatrix} (I_k^q + \frac{1}{8} \Delta t k_2)$$

Here we give an example of how to calculate the integration of first time slot using fourth-order Runge-Kutta with initial $I_k^q = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$. 

$$k_1 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k) \times \rfloor & w(t_k) \\ -w(t_k)^T & 0 \end{bmatrix} I_k^q$$

$$w(t_k + \frac{1}{8} \Delta t) = \frac{1}{2} w(t_k) + \frac{1}{2} w(t_k + \frac{1}{4} \Delta t)$$

$$k_2 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k + \frac{1}{8} \Delta t) \times \rfloor & w(t_k + \frac{1}{8} \Delta t) \\ -w(t_k + \frac{1}{8} \Delta t)^T & 0 \end{bmatrix} (I_k^q + \frac{1}{8} \Delta t k_1)$$

$$k_3 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k + \frac{1}{8} \Delta t) \times \rfloor & w(t_k + \frac{1}{8} \Delta t) \\ -w(t_k + \frac{1}{8} \Delta t)^T & 0 \end{bmatrix} (I_k^q + \frac{1}{8} \Delta t k_2)$$

Here we give an example of how to calculate the integration of first time slot using fourth-order Runge-Kutta with initial $I_k^q = \begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$. 

$$k_1 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k) \times \rfloor & w(t_k) \\ -w(t_k)^T & 0 \end{bmatrix} I_k^q$$

$$w(t_k + \frac{1}{8} \Delta t) = \frac{1}{2} w(t_k) + \frac{1}{2} w(t_k + \frac{1}{4} \Delta t)$$

$$k_2 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k + \frac{1}{8} \Delta t) \times \rfloor & w(t_k + \frac{1}{8} \Delta t) \\ -w(t_k + \frac{1}{8} \Delta t)^T & 0 \end{bmatrix} (I_k^q + \frac{1}{8} \Delta t k_1)$$

$$k_3 = \frac{1}{2} \begin{bmatrix} \lfloor w(t_k + \frac{1}{8} \Delta t) \times \rfloor & w(t_k + \frac{1}{8} \Delta t) \\ -w(t_k + \frac{1}{8} \Delta t)^T & 0 \end{bmatrix} (I_k^q + \frac{1}{8} \Delta t k_2)$$
\[k_4 = \frac{1}{2} \begin{bmatrix} -[w(t_k + \frac{1}{4}\Delta t) \times] & w(t_k + \frac{1}{4}\Delta t) \\ -w(t_k + \frac{1}{4}\Delta t)^T & 0 \end{bmatrix} \begin{bmatrix} I_k^q + \frac{1}{4}\Delta t k_3 \end{bmatrix} \]

\[\Delta q_1 = I_k^q + \frac{\Delta t}{24}(k_1 + 2k_2 + 2k_3 + k_4)\]

Similarly, the remaining three time slots can be calculated and obtain the final integration \(I_{k+1}^q\). The IMU global orientation can be estimated using \(I_{G}^q\) and \(I_k^q\):

\[I_{k+1}^q = I_{k+1}^q \odot I_k^q\] (4.18)

where the \(\odot\) donates quaternion multiplication. The quaternion multiplication of quaternion \(q\) and \(p\) is defined as:

\[q \odot p = \begin{bmatrix} q_4p_1 + q_3p_2 - q_2p_3 + q_1p_4 \\ -q_3p_1 + q_4p_2 + q_1p_3 + q_2p_4 \\ q_2p_1 - q_1p_2 + q_4p_3 + q_3p_4 \\ -q_1p_1 - q_2p_2 - q_3p_3 + q_4p_4 \end{bmatrix}\]

**Accelerometer propagation**

To obtain the velocity and position of the device, we need to represent the acceleration forces with respect to global frame. The acceleration in the global frame can be derived from Eq. 4.14 after bias and gravity compensation:

\[G_a(t) = G_t^R(a_m(t) - b_a(t)) + G_g\] (4.19)
where \( G_I^R \) is the rotational matrix that transform the IMU frame to global frame at time \( t \). The velocity of the device can be derived by integrating Eq. 4.19 and double integration will give the position of the device.

\[
G_{v_{I_{k+1}}} = G_{v_I} + \int_{t_k}^{t_{k+1}} G_{a_I}(\tau) d\tau
\]

\[
= G_{v_I} + \int_{t_k}^{t_{k+1}} G_{I_{r}} R(a_m(\tau) - b_a(\tau)) d\tau + G_{g} \Delta t
\]

\[ \tag{4.20} \]

\[
= G_{v_I} + G_{I_{r}} R s_k + G_{g} \Delta t
\]

where

\[
s_k = \int_{t_k}^{t_{k+1}} I_{r} R(a_m(\tau) - b_a(\tau)) d\tau
\]

\[ \tag{4.21} \]

The position propagation form \( t_k \) to \( t_{k+1} \) can be obtained by integrating Eq. 4.20 again:

\[
G_{p_{I_{k+1}}} = G_{p_I} + \int_{t_k}^{t_{k+1}} G_{v_I}(\tau) d\tau
\]

\[
= G_{p_I} + G_{v_I} \Delta t + G_{I_{r}} R y_k + \frac{1}{2} G_{g} \Delta t^2
\]

\[ \tag{4.22} \]

where

\[
y_k = \int_{t_k}^{t_{k+1}} \int_{t_k}^{t_{k+1}} I_{r} R(a_m(\tau) - b_a(\tau)) d\tau ds
\]

\[ \tag{4.23} \]

Eq. 4.20 and Eq. 4.22 shows how velocity and position are expressed in global frame evolving from time \( t_k \) to time \( t_{k+1} \). Similar to gyroscope propagation, only the sampled version of \( a_m(t) \) at \( t_k \) and \( t_{k+1} \) are available. It is also assumed that the acceleration forces changes linearly from \( t_k \) to \( t_{k+1} \):

\[
a_m(t_k + \frac{1}{4} \Delta t) = \frac{3}{4} a_m(t_k) + \frac{1}{4} a_m(t_{k+1})
\]
\[ a_m(t_k + \frac{1}{2} \Delta t) = \frac{1}{2} a_m(t_k) + \frac{1}{2} a_m(t_{k+1}) \]

Then the integration in Eq. 4.21 and Eq. 4.23 can be computed using Simpson integration [28].

\[ s_k = \frac{\Delta t}{6} (a_m(t_k) + 4 I_{t_k + \frac{1}{2} \Delta t} R a_m(t_k) + \frac{1}{2} \Delta t) + I_{t_k + \frac{1}{2} \Delta t} R a_m(t_{k+1}) \]  

(4.24)

\[ y_k = \frac{\Delta t^2}{18} (a_m(t_k) + 4 I_{t_k + \frac{1}{2} \Delta t} R a_m(t_k) + \frac{1}{4} \Delta t) + I_{t_k + \frac{1}{2} \Delta t} R a_m(t_{k+1}) + \frac{\Delta t}{6} s_k \]  

(4.25)

\( I_{t_k}, I_{t_k + \frac{1}{2} \Delta t}, I_{t_k + \frac{1}{4} \Delta t}, I_{t_k + 1} R \) are rotation matrices which can be obtained from intermediate results (quaternion) during the computation of gyroscope propagation. The conversion from quaternion \( q \) to rotation matrix \( R \) is defined as:

\[
R = \begin{bmatrix}
1 - 2q_2^2 - 2q_3^2 & 2(q_1q_2 + q_3q_4) & 2(q_1q_3 - q_2q_4) \\
2(q_1q_2 - q_3q_4) & 1 - 2q_1^2 - 2q_3^2 & 2(q_2q_3 + q_1q_4) \\
2(q_1q_3 + q_2q_4) & 2(q_2q_3 - q_1q_4) & 1 - 2q_1^2 - 2q_2^2
\end{bmatrix}
\]

4.5.2 Simulation results of IMU propagation

In this section we present the simulation results of the sensor propagation to show how it works. All the simulation data is generated based on real-world datasets, as provided in [30]. The simulation data used is from accelerometer and gyroscope and is verified by ground-truth obtained using stereophotogrammetric system. The sampling rate is 100 Hz and the length of the data is about 2 minutes long. The simulation results are illustrated...
in Fig. 4.8 and Fig. 4.9. It can be seen that as long as the sampling rate is fast enough, the approximation error in Eq. 4.18, Eq. 4.20 and Eq. 4.22 can be negligible using a noise-free IMU sensor.

However, the low-cost built-in IMU sensors exhibit significant systematic errors, axes misalignment and noisy measurement. The systematic errors and axes misalignment can be eliminated through a IMU calibration process. But the noise remains even after calibration and it could cause significant error in Eq. 4.18, Eq. 4.20 and Eq. 4.22 because of
Figure 4.9: Comparison of quaternion propagation using IMU data and quaternion of ground truth.
the integration operation. To demonstrate the effects caused by the IMU noise during the measurements, we add noise to measurements of accelerometer and gyroscope. The noise level in a real IMU “BMI 160” used in Nexus 5X is used as our benchmark. For “BMI 160”, the typical output noise is $180 \mu g/\sqrt{Hz}$ for accelerometer and $0.007^\circ/s/\sqrt{Hz}$ for gyroscope. So we set the continuous time standard deviation of noise as 0.002 in accelerometer and 0.00015 in gyroscope, respectively. For IMU with a 100 Hz sampling rate, this corresponds to a discrete time standard deviation of 0.02 and 0.0015. Fig. 4.10 and Fig. 4.11 shows the propagation of position and quaternion under noise condition. Fig. 4.12 and Fig. 4.13 illustrate the errors in position and quaternion at each time step. It can be seen that the gyroscope is usually more accurate than accelerometer. However, errors in both position and quaternion diverge as the time increases. So additional information is needed to obtain a more accurate result.

4.6 Sensor fusion with extended Kalman filter

4.6.1 Introduction to Kalman filter

In the proposed positioning algorithm, the relative position and pose of the device are kept tracked across the whole positioning process. At each camera frame the device in the global frame are estimated based on the relative position and pose from previous camera frame. This estimation is not accurate as we can see from the simulation. These inaccuracies may include noise from the sensors, approximations in processing stage, the
Figure 4.10: Comparison of position propagation using IMU data and position of ground truth under noise condition.
Figure 4.11: Comparison of quaternion propagation using IMU data and quaternion of ground truth under noise condition.
Figure 4.12: Position error propagation.
Figure 4.13: Rotation angle error propagation.
projection of LED on the image plane, or the temporarily block of the LED optical link. All these measurements are expected to deviate in a random manner from the true value of idealized sensors. All of these inaccuracies result in noise to our positioning process.

To improve the performance of positioning system, we would like to maximize the utilization of the measurements we obtained. Hence, the measurements obtained previously would provide us part of the motion information. The solution is a kinematic model for the motion of device.

This task contains two stages. In prediction stage, information learned in the past are used to evolve the position and pose of the device. In the update stage, measurement is made based on predictions as well as previous measurements. The two-stage task is accomplished using estimators, while Kalman filter is one the most popular estimator [49]. The basic idea of Kalman filter is to model the state of the system that maximize a posterior of previous measurements of the system. Kalman filter develops the system model by combining the system model with uncertainty and new measurement with its own uncertainty, which maximize the probability of estimating correct system states. In a word, Kalman filter corrects the system model by combining previous and new information weighted by their uncertainty.

Considering a physical system model described as:

\[
\dot{x} = f(x, u, w)
\]  

where \(x\) is the state vector, \(u\) is the known system input vector, \(w\) is a zero-mean white
Gaussian process noise with covariance

\[ Q = E[ww^T] \] (4.27)

And the measurement is given by:

\[ z = h(x) + n \] (4.28)

where \( z \) is the measurement vector and \( n \) is the corresponding zero-mean white Gaussian measurement noise with covariance

\[ R = E[ww^T] \] (4.29)

Kalman filter assumes that 1) both \( f(x, u) \) and \( h(x) \) are linear functions; 2) all the noise in the system model are white Gaussian. However, in our application, both the system and measurement are nonlinear functions, therefore, extended Kalman filter (EKF) is employed by linearizing the function using Taylor series expansion at the state estimation \( \hat{x} \). The discrete time form of the EKF is given by the following equations: Propagation: \( \hat{x}_{k+1|k} \) donates the prediction of system state, which is a prior estimate given by:

\[ \hat{x}_{k+1|k} = f(\hat{x}_{k|k}, u(k), 0) \] (4.30)

\( \Phi(k) \) is the error-state transition matrix:

\[ \Phi(k) = \nabla_{x_k} f(x_{k|k}, u(k), 0) \] (4.31)
\( G(k) = \nabla_{w_{k}} f(x_{k|k}, u(k), 0) \) (4.32)

\( P_{k+1|k} \) is the propagated error covariance matrix at time \( k + 1 \) derived from time \( k \).

\[
P_{k+1|k} = \Phi(k)P_{k|k}\Phi^{T}(k) + G(k)Q(k)G^{T}(k)
\] (4.33)

Update: \( \hat{z}_{k+1|k} \) is the estimate of the received measurement:

\[
\hat{z}_{k+1|k} = h(\hat{x}_{k+1|k})
\] (4.34)

The measurement error (residual) is given by:

\[
r_{k+1|k} = z(k + 1) - \hat{z}_{k+1|k}
\] (4.35)

\[
H(k + 1) = \nabla_{x_{k+1}} h(\hat{x}_{k+1|k})
\] (4.36)

The residual covariance can be computed by the following expression:

\[
S_{k+1|k} = H(k + 1)P_{k+1|k}H^{T}(k + 1) + R(k + 1)
\] (4.37)

Then the Kalman gain is computed by

\[
K_{k+1|k} = P_{k+1|k}H^{T}(k + 1)S_{k+1|k}^{-1}
\] (4.38)
Finally, the system state vector and covariance matrix are updated after correction:

\[
\hat{x}_{k+1|k+1} = \hat{x}_{k+1|k} + K_{k+1|k} r_{k+1|k}
\]

(4.39)

\[
P_{k+1|k+1} = P_{k+1|k} - P_{k+1|k} H^T(k+1) S_{k+1|k}^{-1} H(k+1) P_{k+1|k}
\]

(4.40)

### 4.6.2 Implementation of extended Kalman filter

The implementation of EKF is mainly based on the framework developed in [28].

The system state vector at time \( k \) is constructed as:

\[
x_k = \begin{bmatrix}
I^k_G q^T \\
G^k p^T \\
G^k v^T
\end{bmatrix}^T
\]

(4.41)

where \( I^k_G \) is the unit quaternion representing the rotation from global coordinates to IMU coordinates at time \( k \), \( G^k p \) and \( G^k v \) is the IMU position and velocity with respect to global frame at time \( k \).

To derive the state error vector \( \delta x_k \), the orientation error needs to be defined first. The defined orientation error is based on the quaternion \( \delta q \) that expresses the difference between the estimated orientation and the true value:

\[
\delta q = I^k_G \hat{q}^{-1} \otimes I^k_G q
\]

(4.42)

\( \otimes \) represents quaternion multiplication defined in previous section. \( \delta q \) denotes the rotation
from estimation to the true one and if it is small enough it can be written as:

$$\delta q = \begin{bmatrix} \frac{1}{2} G \hat{\Theta}_I \\ 1 \end{bmatrix}$$

(4.43)

with $G \hat{\Theta}_I$ donates the rotational angle errors around three axes in the global frame. Then the state error vector $\tilde{x}_k$ is given by [45]:

$$\tilde{x}_k = \begin{bmatrix} G \hat{\Theta}_I^T \ I_k \\ G \hat{p}_I^T \ I_k \\ G \hat{v}_I^T \ I_k \end{bmatrix}^T$$

(4.44)

where $G \hat{p}_I^T$ and $G \hat{v}_I^T$ are errors of position and velocity with respect to global frame between true values and estimations. Here we directly give the error transition matrix $\Phi_{I_k}$:

$$\Phi_{I_k} = \begin{bmatrix} I_3 & 0_3 & 0_3 \\ -[G I R y_k \times] & I_3 & \Delta t I_3 \\ -[G I R s_k \times] & 0_3 & I_3 \end{bmatrix}$$

(4.45)

where $s_k$ and $y_k$ are defined in Eq. 4.21 and Eq. 4.23. For more details of the derivation of $\Phi_{I_k}$, pleas refer to [28]. After the error transition matrix $\Phi_{I_k}$ is computed, then the measurement Jacobian $H(k)$ can be computed using the following equation:

$$H(k) = \frac{1}{C_{zk}} \begin{bmatrix} 1 & 0 & -\frac{C_{x_k}}{C_{zk}} \\ 0 & 1 & -\frac{C_{y_k}}{C_{zk}} \\ \left[(G p_L - G \hat{p}_I_k) \times\right] & -I_3 & 0_3 \end{bmatrix}$$

(4.46)

where $C_{x_k}, C_{y_k}$ and $C_{z_k}$ are obtained from Eq. 4.11.
4.6.3 Simulation with extended Kalman filter

In this section, the same simulation environment in Sec. 4.5.2 is used except an extended Kalman filter is added to correct the measurement. EKF requires an initial guess of the device position. The initial position is estimated using Eq. 4.7 by setting $Z$ as a default height. The camera frame rate is set as 10 frame per second, which means the Kalman filter will correct 10 times per second. And the standard deviation of LED projection on the image sensor is set as 4 pixels and the standard deviation of initial position is 0.1 m.
Figure 4.15: Orientation propagation corrected by extended Kalman filter (EKF).
Figure 4.16: Position error and the $3\sigma$ bounds.

All the simulation results reported are averaged over 50 Monte Carlo trials.

It can be seen that the maximal error is $0.1941$ m in position and $0.0018$ rad over 50 trials, and all the errors are within the $3\sigma$ bounds.
Figure 4.17: Orientation error and the 3σ bounds.
Chapter 5

Conclusion and Future Work

5.1 Conclusion

In this dissertation, we present a comprehensive analysis of rolling shutter mechanism used in VLC based indoor positioning system. These fundamentals provide guidelines for system design and tool for performance analysis. The shutter speed of the camera is a key parameter that not only determines the observability of rolling shutter effect but also the quality of received signal. Fast shutter speed is preferred but the noise brought in by the increased ISO should be considered. The lower bound of detectable frequency depends on the frame rate, flicker effect as well as the background scene. The upper bound is determined by the exposure time and is usually the order of $1/\tau_{\text{exposure}}$ for reliable detection with good contrast. Higher frequency detection is possible in a low-noise environment. When Fourier spectrum is used to analyze the received signal, DTFT should be used instead of FFT to find a more accurate peak. The frequency resolution of detection is limited by the length and the type of the window used in DTFT analysis. A 14 dB SNR (25 in absolute
value) usually assures a reliable detection with a high probability of signal presence. A background remove technique to compress the unwanted frequency components brought in by the strong ambient light and nonuniform background is proposed. The modulated illumination and background are originally combined by multiplication, not addition. To separate the modulated illumination, a non-linear division operation should be applied with a good approximation of background reflectance. Measurement shows that the proposed background remove algorithm can greatly suppress the interference brought in by the background scene under ambient light environment, especially in the low frequency range close to DC frequency.

Due to the narrow FOV of the camera on the smartphone, usually only one or even none LED could be detected. So these positioning algorithm requiring multiple LEDs would fail under this case. Instead of using multiple LEDs, multiple cameras could be used to calculate the position by employing triangulation. However, there is usually one camera for most smartphone especially for the front facing camera. In our proposed positioning algorithm, we use other built-in sensors of smartphone to provide more information to reduce the required number of LED in the calculation of the position. These sensors not only provide the estimation of the device orientation but also the translation between two camera frames. However, due to the noisy output of these low-cost sensors, the position estimation could quickly diverge from the true value because of the double integration in calculating the system kinematic model. Our proposed approach employing an extended Kalman filter due to the non-linearity of the system. The extended Kalman filter maximize a posterior of previously obtained measurements, by combining the system kinematic prediction and
new measurement of the LED position on the image sensor. A simulation is conducted to verify the proposed positioning algorithm, where the accelerometer and gyroscope are generated from real experiment. The simulation result show that that the position error and orientation error are well bounded in the $3\sigma$ bounds and the maximum position error observed during the 2 minutes simulation is 0.1941 over 50 averaged Monte Carlo trials.

5.2 Future work

The final goal of this work is to implement the positioning algorithm on a smartphone platform. As we discussed in previous section, the low-cost sensors built inside the smartphone often shows significant systematic errors, non-negligible misalignment between sensor axes, different sensitivity along different sensor axes. These problems will not usually occur in high-end sensors. These errors must be modeled and compensated (calibrated) through a real-time process or off-line process. Besides, the employment of rolling shutter mechanism will introduce distortion to the image. The distortion could lead to loss of accuracy in position estimation [26]. When the device exhibits significant accelerations, the readout time of each row must be taken into consideration in order to obtain correct LED position on the image. Another problem in implementation on smartphone platform is the synchronization between inertial sensors and the camera frame [27]. In our simulation, it is assumed that the inertial sensors and camera frame is well synchronized. However, the camera frame and sensor usually use different clock and which is inaccurate. Even the time delay could occur from different types of sensors. These time offset between sensors and camera should be estimated in order to get a more accurate motion estimation. Interpola-
tion could be used to synchronize the time stamps from different sensors and camera.
Bibliography


