# RefRec: Indoor Positioning Using a Camera Recording

# Floor Reflections of Lights

Shota Shimada

Graduate School of
Information Science and Technology
Hokkaido University
Sapporo, Japan
Email: `shimadas@eis.hokudai.ac.jp`

Hiromichi Hashizume

National Institute of Informatics
Tokyo, Japan
Email: `has@nii.ac.jp`

Masanori Sugimoto

Graduate School of
Information Science and Technology
Hokkaido University
Sapporo, Japan
Email: `sugi@ist.hokudai.ac.jp`

*Abstract*—In recent years, there has been a growing interest in indoor positioning techniques using the ubiquitous infrastructure. This paper describes an indoor positioning method using light-emitting diode light reflecting from the floor and a smartphone camera recording it. Almost all traditional methods must detect the light source directly. However, there are some constraints to their usage since light sources cannot always be detected directly. Our proposal aims to solve this problem by estimating the position of a camera that does not face the light directly but records light reflected from the floor. The camera need not detect the ceiling lights directly from an image, unlike existing methods. Experimental results show that the proposal requires less than 1/100 the number of pixels for localization as do existing methods, and the 3-D position and attitude can be estimated within 0.27 m and 5.78 degrees at the 90th percentile in a 4.0 m square room.

*Keywords–Visible light positioning; Received signal strength; Angle of arrival.*

## I. INTRODUCTION

With the development of ubiquitous computing, there is great interest in Indoor Positioning and Indoor Navigation (IPIN) technology for mobile devices. In 2017, United States (US)\$2,642 million in revenue from IPIN techniques was realized and this revenue is predicted to reach US\$43,511 million by 2025 [1]. That is because IPIN has valuable applications in medical care, manufacturing, advertising, and sales, amongst others. In particular, the retail industry led the IPIN market in 2017 and predicted that demand such as improved customer searches, effective route planning, and optimized customer targeting would continue to increase. Customers usually do not prefer to obtain additional devices for IPIN, whereas many people have mobile phone devices.

From this background, many methods of the indoor positioning for mobile devices have been proposed. These utilize radio waves, infrared, sound, computer vision, visible lights, and others [2]. Among them, Visible Light Positioning (VLP), using a transmitter and receiver constructed by Light-Emitting Diode (LED) and camera or Photodiode (PD), has shown some promise for indoor positioning [3]. The receiver recognizes the light of the transmitters and then calculates its relative position to the transmitters. Compared with other methods such as those based on radio waves, VLP has three major advantages. First, VLP is efficient because LEDs can be used not only as lighting for humans but also as transmitters for positioning. Second,

visible light does not penetrate objects, which reduces the multipath problem. Third, it is easier to install the VLP system because many buildings have the infrastructure for lighting.

Usually, LEDs should be modulated in high-frequency ranges to avoid humans sensing flickering, and to enable detection by the receiver. However, PDs used in Commercial Off-The-Shelf (COTS) mobile devices are not sensitive enough and their response rates are too slow for VLP. Therefore, we focus on a camera-based method using only a mobile device as a receiver.

Most of the existing camera-based methods use a large image for positioning because they need to detect multiple light sources installed at different places on a ceiling, directly from the image. It is computationally demanding, and some existing methods cannot calculate in real time using smartphones [4][5]. Also, it is too difficult to capture several LEDs into the image if the ceiling is low (it means a Loss of Signal: LOS). The user will therefore be forced to detect the LEDs by moving the camera. Some prior work also requires changing the camera to use a Complementary Metal Oxide Semiconductor (CMOS) image sensor that implements a rolling shutter. This feature enables multiple samplings in one image, but customers usually do not like the distortion of the image caused by the rolling shutter. However, there are developments to overcome this problem and this effect may disappear in the future [6].

In order to address these problems, we propose the VLP method RefRec, which does not find a light source directly but records reflected light on the floor. In our proposal, a camera captures light from the ceiling that is reflected by the floor and estimates the distance from the LED. No matter where the camera captures the floor reflection, the ceiling light will be reflected everywhere on the floor. The challenge is that reflections of multiple lights will be overlapped. To separate them, the frame rate of the camera and each frequency of LED light is determined by DC-biased Optical-Orthogonal Frequency Division Multiplexing (DCO-OFDM) [10]. Our experimental in a 4.0 m square room shows that our method requires $32 \times 32 \times 9$ pixels for self-positioning within 0.27 m and $5.78°$ at the 90th percentile. A comparison with existing methods is shown in Table I. The details are described in Section II.

Our contributions are summarized as follows:

TABLE I. COMPARISON WITH EXISTING VLP METHODS USING LED.

| | Epsilon [7] | Luxapose [4] | PIXEL [8] | Rajagopal [5] | Nakazawa [9] | Our proposal |
|---|---|---|---|---|---|---|
| **Accuracy of position** | $\sim$0.4m | $\sim$0.1m | $\sim$0.3m | N/A | $\sim$0.1m | $\sim$0.27m |
| **Accuracy of rotation** | N/A | $\sim 3^\circ$ | N/A | N/A | N/A | $\sim 5.78^\circ$ |
| **Method** | Model | AoA | Polarized | PRR | Model | RSS |
| **Coverage** | 5.0 m×8.0 m | 1.0 m×1.0 m | 2.4 m×1.8 m | 3.9 m×8.0 m | 1.0 m×2.4 m | 4.0 m×4.0 m |
| **Facing** | Ceiling | Ceiling | Ceiling | Floor | Ceiling and floor | Floor |
| **Distortion** | No | Yes | No | Yes | Yes | No |
| **Extra** | PD | No | Filter | No | No | No |
| **Number of LEDs** | 5 | 5 | 8 | 4 | 2 | 4 |
| **Resolution** | N/A | 7712×5360 | 120×160 | 1280×720 | 3280×2460 1472×1104 | 32×32×9 |

- The VLP method does not require large images and does not have rolling shutter distortion.
- The six-degrees-of-freedom mobile attitude estimation algorithm uses only modulated LEDs and a camera.
- Performance evaluation of the proposal was carried out by real-time positioning experiments.

We describe the challenges of prior works in Section II, our new method for six-degrees-of-freedom localization in Section III, the details of our prototype in Section IV, the experiments to demonstrate the advantages of our method in Section V, and the limitations of our proposal in Section VI. Our conclusion and future work are summarized in Section VII.

## II. RELATED WORK

To explain the VLP method using PDs or cameras, some existing methods are selected, and their challenges are discussed.

### A. Visible light positioning using PDs

Among the methods using PDs, Time of Arrival (ToA) [11], Time Difference of Arrival (TDoA) [12], Angle of Arrival (AoA) [13], and Received Signal Strength (RSS)-based [14] ones have been proposed. Our proposed method utilizes these previous RSS-based studies. Epsilon was the first indoor VLP designed in the academic community [7]. It detects the binary shift keying of the LED using the prototype device with the PD and derives the position by triangulation. Accuracies of 0.4 m, 0.7 m, and 0.8 m at the 90th percentile were achieved in three different office spaces. NALoc uses the same type of device as the ambient light sensor that is implemented in a smartphone. The results gave 90th-percentile errors of less than 0.35 m for the 2-D position, but the device was evaluated separately from the smartphone and not in a built-in setting [15].

### B. Visible light positioning using camera

Camera-based methods allow geometrical separation of the light sources, allowing for more accurate positioning [16][17]. Luxapose can compute the position and posture of the smartphone by capturing ceiling lights directly with a camera [4]. The error is less than 10 cm and less than 3°. It uses 7712×5360 pixels in a WindowsPhone 8 smartphone camera as the receiver. The calculation requires a cloud server for high-quality image processing. PIXEL is a polarization-based localization method [8]. Only 120 × 160 pixels are required. It can be measured in several seconds with an accuracy of 0.4 m. However, a polarizing filter must be attached to the camera, so there is a risk of impairing the original image. Rajagopal's approach uses light reflected by the floor [5]. It

is similar to our idea, but they focus on the rolling shutter distortion to receive an Identifier (ID) from the reflected light. Carriers up to 8 kHz can be received with a channel separation of 200 Hz. Tag information is transmitted by assigning ON and OFF bits to different frequencies. The data rate is 10 bps, and up to 29 light sources can be uniquely separated. Positioning accuracy was not discussed in the paper because the research aimed for the semantic positioning from differences in Packet Reception Rates (PRR). Further, because MATLAB was used for calculation, processing was not in real time. Nakazawa's method uses a dual-facing camera and calculates its own position by the relationship between ceiling light and reflected light on the floor. Large coverage and high accuracy are achieved using only two LEDs. This method also requires a large image so average processing time is 1.2 sec [9]. A comparison of conventional methods is summarized in Table I. Distortion means dose these methods require the rolling shutter distortion or not. Extra indicates an additional device for VLP.

## III. SYSTEM DESCRIPTION

An overview of our system is shown in Figure 1. Our method is based on the following key approaches. RSS from the LED will decrease with the distance from the LED to the floor. More than two LEDs are mounted on the ceiling, which is parallel to the $X$–$Y$ plane. The $k$th LED's known 3-D coordinate is $(x_{L_k}, y_{L_k}, z_h)$. Each LED broadcasts a sinusoid wave with its own unique frequency. Modulated light from the LEDs is reflected by the floor, which is parallel to the $X$–$Y$ plane where $Z = 0$. The camera of the mobile device captures any part of the floor $P_{F_i}$, then estimates its own position and attitude by means of three steps.

First, each distance $d_{i,k} = \sqrt{(x_{L_k} - x_{F_i})^2 + (y_{L_k} - y_{F_i})^2}$ is estimated. Second, the positions of several points $P_{F_i}$ on the floor captured by the camera are estimated using $d_{i,k}$. Finally, the camera position $(x, y, z)$ and attitude $(\theta_x, \theta_y, \theta_z)$ are estimated by optical AoA, using each $P_{F_i}$. The definition of attitude $(\theta_x, \theta_y, \theta_z)$ for a smartphone is shown in Figure 2a. The head of the smartphone points in the $Y_c$ direction and the camera faces the floor at its initial status $(\theta_x, \theta_y, \theta_z) = (0, 0, 0)$. Also, we define 2-D coordinates in the image as shown in Figure 2b. The following sections provide details of the three steps.

### A. Distance between the light source and the photographed floor

The first step is to find the distance $d_{i,k}$ from the intersection $P_k$ of the perpendicular passing through the $k$th light source and the floor to the point $P_{F_i}$ captured by the camera. The camera detects the intensity of the received signal from the
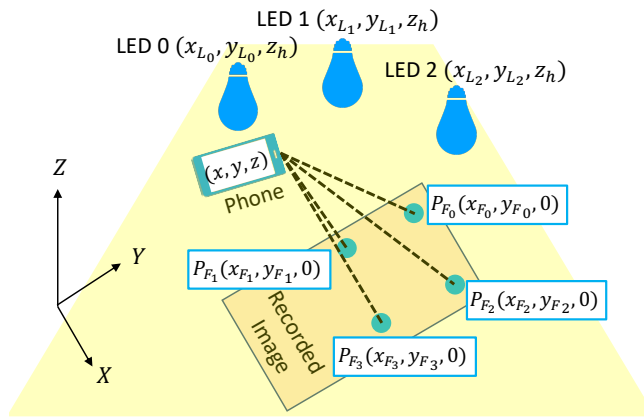
Figure 1. System overview.



(a) Definition of attitude for a smartphone.
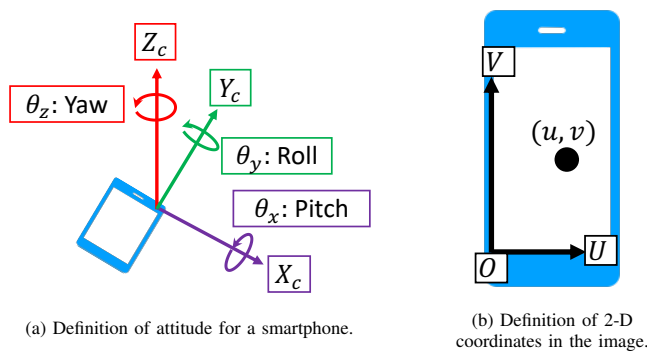
(b) Definition of 2-D coordinates in the image.

Figure 2. Definitions of attitude and 2-D coordinates for mobile positioning.

photographed image and substitutes it into the diffusion model of the LED reflected light. To reduce the influence of other light sources, the LED blinks a sinusoidal wave orthogonal in frequency to those of other light sources. We define $b_k(t)$, which includes a signal from the $k$th LED at time $t$, as follows:

$$b_k(t) = \sin(2\pi t(A_k f_s + m_k)) + \alpha, \quad (1)$$

where $f_s$ is the basic signal frequency and $A_k$ is a natural number so that the signal gets to a high enough frequency not to cause flickering; $\alpha$ is the direct current component, which makes $b_k(t)$ always a positive value; and $m_k$ is a natural number that uniquely identifies the frequency for each LED.

Assuming that the camera frame rate is $f_c = f_s$, the shutter cycle is $T_c = 1/f_c$, the exposure time ratio is $\eta$, and the exposure time is $\eta T_c$. By taking $N$ images with the camera and separating the received light in the frequency domain, it is possible to extract the signal intensities of the unique frequencies. Therefore, the number of detectable LEDs is $N/2 - 1$, and $m_k < N/2$ should be satisfied as per the sampling theorem. By capturing $b_k(t)$ from the $k$th LED with a camera, the resulting brightness $I_{i,n}$ of the $P_{F_i}$ on the $n$th image is an integral:

$$I_{i,n} = \sum_k \frac{2\pi X(d_{i,k})}{T_c} \int_0^{\eta T_c} b_k(t + \delta T_c + nT_c)dt. \quad (2)$$

where $\delta$ is the delay of the shutter timing with respect to the signal, and $X(d_{i,k})$ is the attenuation function that is determined by the distance and transfer efficiency from the $k$th LED as a transmitter to the receiver. Hence, our purpose is to calculate $d_{i,k}$ from the inverse function of $X(d_{i,k})$ using $I_{i,n}$.

Now, $B_{i,l}$ is obtained by the Fourier transform of the video stream $\mathbf{I}_i = (I_{i,0}, I_{i,1}, ..., I_{i,N-1})$:

$$B_{i,l} = \frac{1}{N} \sum_{n=0}^{N-1} I_{i,n} e^{\frac{-j2\pi nl}{N}}. \quad (3)$$

By assuming $\beta_{(m_k,k)}$ obtained by Fourier transform of $b_k(t)$, $B_{i,l}$ is shown as follows [18]:

$$B_{i,m_k} = \eta X(d_{i,k}) e^{j\pi f_k(2\delta+\eta)} \text{sinc}(\pi f_k \eta) \beta_{(m_k,k)} \quad (4)$$

where $f_k = A_k f_s + m_k$. Thus, the unique frequency $m_k$ can be extracted. Note that the amplitude spectrum $|\beta_{(m_k,k)}|$ is affected by $\eta$ and sinc. Attenuation of magnitude affects the accuracy of positioning, so $\eta$ must be set uniquely. Note that $X(d_{i,k})$ is the product of attenuation of the LED's signal and transfer efficiency. We assume the attenuation of the LED's signal is inversely proportional to the square of the distance and attenuates by the cosine of the radiation angle $\theta$. However, it is difficult to model the attenuations theoretically because reflecting properties are very complicated in the real environment [19]. Our previous work showed attenuation on the floor can be approximated by a hyperbolic secant distribution [20]. Therefore,

$$X(d_{i,k}) = \frac{C_k}{e^{\frac{\pi}{2}\sigma d_{i,k}} + e^{-\frac{\pi}{2}\sigma d_{i,k}}}, \quad (5)$$

where $\sigma$ is the radiation characteristic of the LED, and $C_k$ is the transmission efficiency determined by the receiver sensitivity. Hence, the amplitude spectrum $|B_{i,m_k}|$ is shown as follows:

$$|B_{i,m_k}| = \eta X(d_{i,k}) \text{sinc}(\pi f_k \eta) |\beta_{(m_k,k)}|. \quad (6)$$

Thus, $d_{i,k}$ can be calculated as

$$d_{i,k} = \frac{1}{\sigma} \cosh^{-1}\left(\frac{C'_k}{|B_{i,m_k}|}\right) \quad (7)$$

where $C'_k = \eta C_k \text{sinc}(\pi f_k \eta) |\beta_{(m_k,k)}|$.

### B. Position on the floor recorded by a camera

The second step is to obtain $P_{F_i}$-captured positions on the floor, using three or more $d_{i,k}$ estimated by the previous step. Now, we assume that three perpendiculars from the light sources pass through the floor at the points $(x_{L_0}, y_{L_0})$, $(x_{L_1}, y_{L_1})$, and $(x_{L_2}, y_{L_2})$ (see Figure 1). The estimated distances of these intersections to the point $(x_{F_i}, y_{F_i})$ are $d_{i,0}$, $d_{i,1}$, and $d_{i,2}$, respectively:

$$\begin{cases} \sqrt{(x_{F_i} - x_{L_0})^2 + (y_{F_i} - y_{L_0})^2} = d_{i,0} \\ \sqrt{(x_{F_i} - x_{L_1})^2 + (y_{F_i} - y_{L_1})^2} = d_{i,1} \\ \sqrt{(x_{F_i} - x_{L_2})^2 + (y_{F_i} - y_{L_2})^2} = d_{i,2} \end{cases} \quad (8)$$

By solving this, $(x_{F_i}, y_{F_i})$ can be calculated. When there are more than three LEDs installed in the building, it can be solved as an optimization problem:

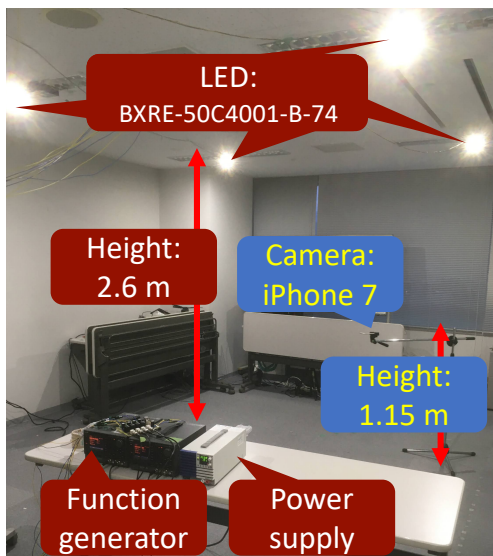$$\min \sum_k \left(\sqrt{(x_{F_i} - x_{L_k})^2 + (y_{F_i} - y_{L_k})^2} - d_{i,k}\right)^2. \quad (9)$$

Figure 3. Positioning of the LEDs in the room and the camera settings.



Figure 4. Absolute mean error of estimated distance for each ISO and shutter speed setting.

### C. Position and attitude of the camera

The last step is to estimate the position and attitude of the camera. These are calculated using the optical AoA method with the traditional camera matrix

$$sp = M[R \ \mathbf{t}]P, \qquad (10)$$

where $s$ means the scale coefficient, $M$ denotes the intrinsic properties, and $[R \ \mathbf{t}]$ represents the extrinsic properties [21]. $p$ is the 2-D image coordinates, and $P$ defines the 3-D world coordinates as follows:

$$p = \begin{bmatrix} u_{I_0} & ... & u_{I_i} & ... \\ v_{I_0} & ... & v_{I_i} & ... \\ 1 & ... & 1 & ... \end{bmatrix} \qquad (11)$$

$$P = \begin{bmatrix} x_{F_0} & ... & x_{F_i} & ... \\ y_{F_0} & ... & y_{F_i} & ... \\ 0 & ... & 0 & ... \\ 1 & ... & 1 & ... \end{bmatrix} \qquad (12)$$

where the 3-D world coordinates $(x_{F_i}, y_{F_i})$ are captured by a camera as 2-D image coordinates $(u_{I_0}, v_{I_0})$ on the image. The camera position and rotation matrix $[R \ \mathbf{t}]$ is obtained by minimization of $||\mathbf{A}[R| \ \mathbf{t}]P - sp||_2$.

### IV. IMPLEMENTATION DETAILS

The prototype of RefRec using LEDs and a smartphone was implemented in our laboratory as shown in Figure 3. A floor made from a patternless nonglow mat was chosen because our prior work has shown that floor material might affect the result adversely [20].

### A. LED transmitter

BXRE-50C4001-B-74-type LEDs from Bridgelux were used as the LED transmitter. Our proposal assumes that the light source is not an area or a line source, e.g., a flat panel or a bar light. Extending the experiments to including these sources remains to be done in future work. Four LEDs above the room were set as shown in Figure 3. The height of the ceiling is 2.6
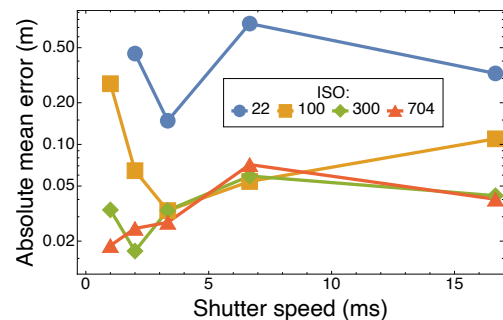
m, coverage is 4.0 m square, LEDs 0, 1, 2, and 3 were set at $(1.0, 0.5, 2.6)$, $(1.0, 3.0, 2.6)$, $(3.2, 0.5, 2.6)$, and $(3.2, 3.0, 2.6)$ m. This setting was the same as the original built-in formation of fluorescent lights in this room. No other objects were placed in the room to evaluate the performance properly in an ideal environment. The signal parameters were set at $f_s = 50$, $\alpha = 1$, $A_k = 2$ (at any $k$), and $m_k = (1, 6, 13, 20)$. Hence, the LEDs were modulated at 101, 106, 113, and 120 Hz. These frequencies are higher than the 100 Hz modulation frequency for fluorescent lights in east Japan to ensure people do not experience any flickering. The transmitter represented any signal by the Pulse Density Modulation (PDM). A 5 V pulse signal from a function generator (NF Corporation WF-1948) was amplified to 34 V using a Metal-Oxide-Semiconductor Field-Effect Transistor (MOS-FET) K703 and a power supply. The frequency of the pulse signal was about 8 MHz. This was a prototype and the transmitter can be made cheaper and smaller by using a circuit similar to a dimmable off-the-shelf LED in our future experiments.

### B. Camera receiver

An iPhone 7 was used as a receiver. Smartphones in recent years have generally higher-performance chipsets and cameras than the iPhone 7. Therefore, we believe that the experiments in this paper can be reproduced on other smartphones that users have. The frame rate $f_c = N = 50$ was set to avoid effects from other fluorescent lights. In east Japan, fluorescent lights are modulated by AC 100 Hz, so 50 fps is the orthogonal frequency, which will treat other fluorescent lights just as DC sources (the same as sunshine). The F value is fixed on f/1.8 in the case of the iPhone 7. The shutter speed and ISO sensitivity should be arbitrarily set so that the pixels are not saturated. The focus was fixed on the floor. Distortion of the image was calibrated using Zhang's method [22].

To reveal the relationship between the camera parameter and positioning accuracy, for each ISO and shutter speed setting, each distance was estimated by the following equation:

$$d_0 = \sqrt{(x_{L_0} - x_F)^2 + (y_{L_0} - y_F)^2}. \qquad (13)$$

Note that $(x_F, y_F)$ is the floor 2-D coordinates $P_F$ captured by the principal point on the image. The iPhone 7 was fixed horizontally to a tripod 1.15 m above the floor and moved from $d_0 = 0$ m to $d_0 = 3.5$ m in 0.5 m increments. Each distance was estimated 100 times and absolute mean errors are shown in Figure 4. When the ISO value was set too low, the distance $d_0$ could not be estimated correctly. When the ISO value was set

over 300, estimates at the centimeter level were achieved. We also found that shutter speeds should be shortened. However, shutter speeds that are too short make estimation difficult because the images become too dark. Hence, the settings of the camera are suggested that the ISO value is set at over 300, and the shutter speed is set at less than 1/300 sec. We subsequently set the shutter speed at 1/500 sec. and the ISO value at 500.

## V. EVALUATION

In order to clarify the advantages and limitations of our proposal, the performance of RefRec was evaluated in our experiment.

### A. Estimation of captured floor $P_F$ positions

First, one $P_F$ estimated by equation (9). The iPhone 7 does not support small resolution read-outs, so the resolution is set at 960×540 pixels, and used only 32×32 pixels around the principal point. Our previous work revealed this resolution to be smaller than that used in conventional methods and it enables real-time performance, while sufficient accuracy could be achieved for many applications [20]. The arrangement of the $P_F$s are shown in Figure 5a as filled circles. The triangles indicate the LED positions on the ceiling. Since the parameters of equation (7) are affected by the individual differences of smartphone cameras and LEDs, calibration is performed before evaluation. Specifically, a camera was set with a tripod below LED 0 to receive a signal, and then updated the constant $C'_k$ in equation (7) with $d_0 = 0$. The position of each $P_F$ was estimated 100 times. All measurements were performed in real time on the smartphone. The mean positions of the $P_F$s estimated by the camera are shown as non-filled circles in Figure 5a and the errors from the true $P_F$ positions are shown as arrows. The error became larger outside the rectangular area with the coordinates of the four LEDs as vertices. Therefore, in order to measure a wider area, it is necessary to arrange more LEDs to increase the rectangular area. Since the LED beacons were set as square, the error should ideally be point symmetric with $(x, y) = (2.1, 1.75)$ as the origin coordinates. However, Figure 5a did not show symmetry because each LED has a slightly different feature. The cumulative distribution function of the estimated absolute errors are shown in Figure 5b. An estimation error of less than 0.42 m at the 90th percentile was achieved.

### B. Estimation of mobile device attitude using captured floor $P_F$ positions

Next, nine $P_{F_i}$ $(0 \le i < 9)$ were estimated by one image stream. A resolution of 1920×1080 pixels was set. The principal point was $(u_c, v_c) = (524.86, 959.07)$. As Figure 6a shows, the 2-D image coordinates in $p$ are $(u_{I_j}, v_{I_k}) = (300j + 200 - u_c, 300k + 200 - v_c)$ $(0 \le i < 3, 0 \le j < 3)$. A camera was set in the room with parameters $(x, y, z) = (1.0, 0.5, 1.15)$ and $(\theta_x, \theta_y, \theta_z) = (0, 0, \pi/4)$. Each $P_F$ position was estimated 100 times in real time and the positions are shown in Figure 6b. The accuracy of the estimated azimuth $\theta_z$ for every combination of $P_{F_i}$ was evaluated. The total number of $P_{F_i}$s is nine, so the total number of combinations of choosing more than one point from these is 502. This evaluation was processed off-line. The estimated absolute median error at each total number of $P_{F_i}$s used to calculate $\theta_z$ is shown in Figure 7a. When the number of $P_{F_i}$s used is two, the estimated errors are very different from



(a) LED arrangement, $P_F$ measurement points, and estimated positions.



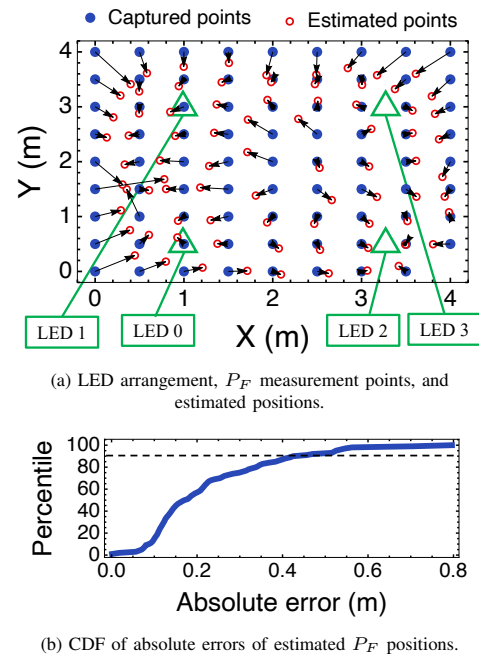(b) CDF of absolute errors of estimated $P_F$ positions.

Figure 5. LED arrangement and estimated errors of $P_F$ positions.

which the $P_{F_i}$s are selected. The best 90th percentile absolute error was $6.15°$, with the worst 90th percentile absolute error at $27.09°$. When the baseline length was long (e.g., $(u_{I_0}, v_{I_0})$ and $(u_{I_2}, v_{I_2})$), the accuracy improved. On the contrary, the results for the cases with short baseline lengths (e.g., $(u_{I_1}, v_{I_2})$ and $(u_{I_2}, v_{I_2})$) were inaccurate. As the number of $P_{F_i}$s increased, the difference between combinations (standard deviation) decreased. When the number of $P_{F_i}$s used was eight, the estimated errors were similar, with the best 90th percentile absolute error of $3.25°$, and the worst 90th percentile absolute error of $5.77°$. In the next section, we discuss the use of all the $P_{F_i}$-captured positions.

The six-degrees-of-freedom mobile device attitude were estimated using nine obtained $P_{F_i}$s. The results are shown in Figure 7. The estimated 3-D coordinate positions of the smartphone $(x, y, z)$ are shown in Figure 7b. The 90th percentile absolute errors of the $x$, $y$, and $z$ coordinates are 0.2073 m, 0.1713 m, and 0.002464 m. Thus, an absolute error of less than 0.27 m in 3-D localization was achieved. $x$ and $y$ are more sensitive than $z$. The estimated mobile attitude $(\theta_x, \theta_y)$ is shown in Figure 7c. The 90th-percentile attitude errors were less than $2.47°$ and $4.93°$ for the pitch and roll angles. The cumulative distribution function for the azimuth Z in Figure 7d shows a 90th percentile absolute error is less than $3.45°$. The pitch and roll of the smartphone can also be obtained precisely using the Inertial Measurement Unit (IMU) [23]. However, it is difficult to estimate the azimuth using IMU because the magnetic field is unstable in the room. Our results show RefRec has potential for more mobile applications.

### C. Angle precision of different attitude of the smartphone

To investigate the accuracy limitations of the smartphone's angle estimation, orientation errors from $P_{F_i}$ were evaluated for each attitude of the smartphone. The pitch, roll, and yaw of the smartphone were changed by 10 degrees each, and its
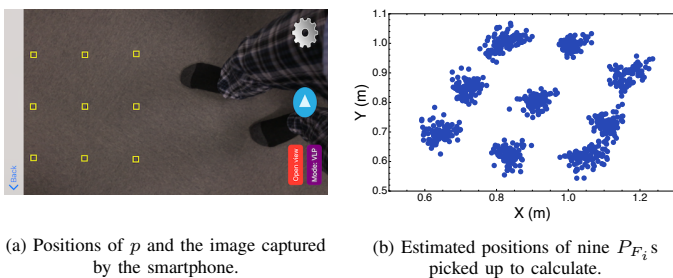
(a) Positions of $p$ and the image captured by the smartphone.

(b) Estimated positions of nine $P_{F_i}$s picked up to calculate.

Figure 6. Arrangement of estimated $P_{F_i}$s and how many $P_{F_i}$s affect positioning error.



(a) Azimuth $\theta_z$ at each total number of points $P_{F_i}$ used.

(b) 3-D coordinates of mobile position $x$, $y$, and $z$.

(c) Pitch $\theta_x$ and roll $\theta_y$.

(d) Azimuth (yaw) $\theta_z$.
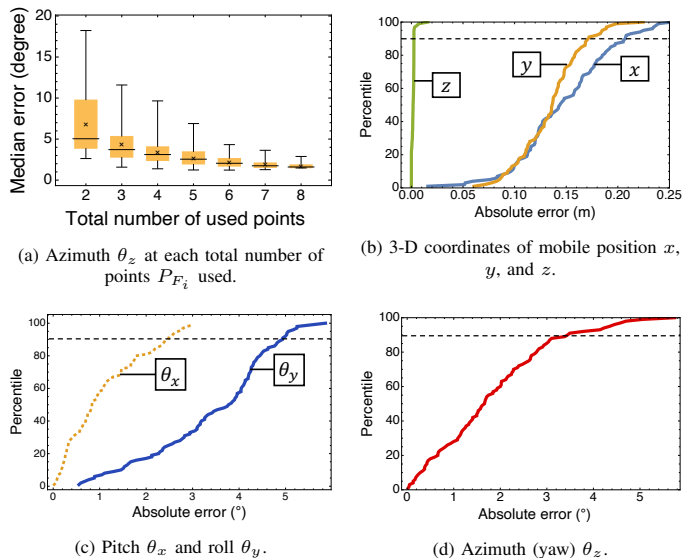
Figure 7. CDF of estimated absolute errors about six degrees-of-freedom using nine $P_{F_i}$s.

attitude was estimated 100 times. Considering the use case of holding a smartphone, the roll, pitch, and yaw were limited to $-30°$ to $30°$, $0°$ to $30°$, and $0°$ to $90°$. Each 90th percentile absolute rotation error at each posture of the smartphone is shown in Figure 8. Means of 90th percentile absolute rotation error were $5.69°$, $5.78°$, and $3.96°$ for the roll, pitch, and yaw.

## VI. DISCUSSION

In this section, we discuss the limitations of our proposal and potential remaining future work.

### A. Comparison of performance

A comparison of the performance of positioning is presented in Table I. Please note that each experimental environment differs in terms of LED installation spacing and measurement area size. Performance comparisons in the same condition are difficult to make because these differences have a significant impact on accuracy. For example, the best-accuracy Luxapose in Table I is difficult to localize in our experimental environment where the LED spacing is too wide to record multiple LEDs directly using smartphone camera.
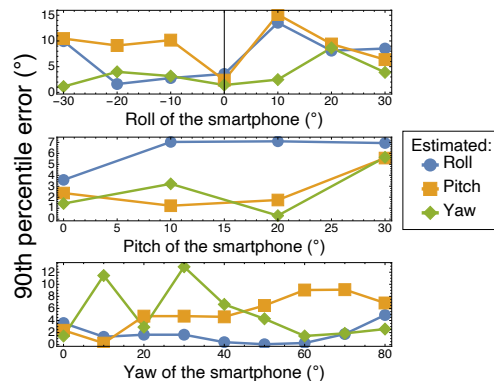


Figure 8. 90th percentile absolute rotation error at each posture of the smartphone.

### B. Materials of transmitter and receiver

Changing the materials of the floor and the LEDs as transmitters may cause errors. In particular, LEDs with a Lambertian emission radiation characteristic are better. Our previous research showed that the difference of reflections resulting from different floor materials causes errors [20]. Future work must include investigating floors that consist of glowing materials or contain some patterns.

### C. Attitude of smartphone

The attitude of the smartphone, especially its pitch and roll (tilt) attitude, may cause errors. We suggest that the smartphone should be held horizontally. We are of the opinion that this is easier to achieve than the existing method, which requires moving the mobile to seek the LEDs directly.

### D. Power consumption

The camera consumes power using our positioning method. Approximately 300 mW will be consumed using an off-the-shelf camera [24]. If the user wants to navigate, it is necessary to continue activating the camera, so it consumes a considerable amount of power. However, because only a few pixels are required by our method, the power requirement can be reduced if the image sensor can activate only the necessary elements.

### E. Solution in the real environment

In our experiments, nothing was placed in the measurement space in order to minimize signal changes. However, in a real environment there are many obstacles, such as objects or humans. The shadow of so-called ambient occlusion might be harmful to positioning. In the future, we will propose how to select an area that does not include shadows. We will also investigate influences and conduct experiments in larger spaces using, for example, 16 LEDs in a 10 m square room.

## VII. CONCLUSION AND FUTURE WORK

Visible light positioning for smartphones is regarded as a promising technique that is expanding the market in many industries. This paper describes an approach to avoid existing limitations, such as LOS, by using a camera recording light reflected by the floor. Our prototype showed 90th percentile 3-D localization and attitude estimation errors that were within

0.27 m and $5.78°$. Our proposal has larger coverage and a smaller image requirement than conventional techniques. We also mention that some conditions may affect the positioning accuracy, such as the floor materials, radiation characteristics of the LED, and tilt of the smartphone, amongst others. A combination of our proposal and conventional techniques should reduce the limitations and improve accuracy. Future work will explore cases where, for example, more people hold their smartphones and move around, objects are placed on the floor to cause occlusion, more LEDs are used, and experimenting is done in a larger area.

### ACKNOWLEDGMENT

### REFERENCES

[1] P. Lanjudkar, "Indoor Positioning and Indoor Navigation (IPIN) Market," https://www.alliedmarketresearch.com/indoor-positioning-and-indoor-navigation-ipin-market, 2018, [retrieved: Sep,2020].

[2] P. Davidson and R. Piché, "A survey of selected indoor positioning methods for smartphones," IEEE Communications Surveys & Tutorials, vol. 19, no. 2, 2016, pp. 1347–1370.

[3] S. D. Lausnay, L. D. Strycker, J. P. Goemaere, B. Nauwelaers, and N. Stevens, "A survey on multiple access visible light positioning," in Proc. 2016 IEEE Int. Conf. on Emerging Technologies and Innovative Business Practices for the Transformation of Societies, Port Louis, Mauritius, Aug 2016, pp. 38–42.

[4] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," in Proc. 20th annual Int. Conf. Mobile Computing and Networking, Maui, Hawaii, 2014, pp. 447–458.

[5] N. Rajagopal, P. Lazik, and A. Rowe, "Visual light landmarks for mobile devices," in Proc. 13th Int. Symp. Information Processing in Sensor Networks, Berlin, Germany, 2014, pp. 249–260.

[6] Sony Semiconductor Solutions Corporation, "Sony Develops the Industry's First*1 3-Layer Stacked CMOS Image Sensor with DRAM for Smartphones," https://www.sony.net/SonyInfo/News/Press/201702/17-013E/, [retrieved: Sep,2020].

[7] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao, "Epsilon: A visible light based positioning system." in Proc. 11th USENIX Symp. Networked Systems Design and Implementation, vol. 14, Seattle, WA, United States, 2014, pp. 331–343.

[8] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang, "Wearables can afford: Light-weight indoor positioning with visible light," in Proc. 13th Annual Int. Conf. Mobile Systems, Applications, and Services, New York, NY, United States, 2015, pp. 317–330.

[9] Y. Nakazawa et al., "Precise indoor localization method using dual-facing cameras on a smart device via visible light communication," IEICE Trans. Fundamentals of Electronics, Communications and Computer Sciences, vol. E100.A, no. 11, 2017, pp. 2295–2303.

[10] J. Armstrong and A. Lowery, "Power efficient optical ofdm," Electronics letters, vol. 42, no. 6, 2006, pp. 370–372.

[11] T. Q. Wang, Y. A. Sekercioglu, A. Neild, and J. Armstrong, "Position accuracy of time-of-arrival based ranging using visible light with application in indoor localization systems," Journal of Lightwave Technology, vol. 31, no. 20, 2013, pp. 3302–3308.

[12] S.-Y. Jung, S. Hann, and C.-S. Park, "Tdoa-based optical wireless indoor localization using led ceiling lamps," IEEE Trans. Consumer Electronics, vol. 57, no. 4, 2011.

[13] S.-H. Yang, H.-S. Kim, Y.-H. Son, and S.-K. Han, "Three-dimensional visible light indoor localization using aoa and rss with multiple optical receivers," Journal of Lightwave Technology, vol. 32, no. 14, 2014, pp. 2480–2485.

[14] H. Steendam, T. Q. Wang, and J. Armstrong, "Theoretical lower bound for indoor visible light positioning using received signal strength measurements and an aperture-based receiver," Journal of Lightwave Technology, vol. 35, no. 2, Jan 2017, pp. 309–319.

[15] L. Yang, Z. Wang, W. Wang, and Q. Zhang, "Naloc: Nonlinear ambient-light-sensor-based localization system," Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 2, no. 4, 2018, pp. 1–22.

[16] M. S. Rahman, M. M. Haque, and K.-D. Kim, "Indoor positioning by led visible light communication and image sensors," International Journal of Electrical and Computer Engineering, vol. 1, no. 2, 2011, p. 161.

[17] M. Yoshino, S. Haruyama, and M. Nakagawa, "High-accuracy positioning system using visible led lights and image sensor," in Proc. Radio and Wireless Symposium, Orlando, FL, United States, 2008, pp. 439–442.

[18] S. Shimada, T. Akiyama, H. Hashizume, and M. Sugimoto, "Ofdm visible light communication using off-the-shelf video camera," in Proc. 15th ACM Conf. Embedded Network Sensor Systems, Delft, Netherlands, 2017, p. 57.

[19] H. Zhang and F. Yang, "Push the limit of light-to-camera communication," IEEE Access, vol. 8, 2020, pp. 55 969–55 979.

[20] S. Shimada, H. Hashizume, and M. Sugimoto, "Indoor positioning using reflected light and a video camera," in Proc. 9th Int. Conf. Indoor Positioning and Indoor Navigation, Nantes, France, 2018, pp. 1–8.

[21] G. Bradski and A. Kaehler, Learning OpenCV: Computer vision with the OpenCV library. " O'Reilly Media, Inc.", 2008.

[22] Z. Zhang, "A flexible new technique for camera calibration," IEEE Trans. pattern analysis and machine intelligence, vol. 22, no. 11, 2000, pp. 1330–1334.

[23] P. Zhou, M. Li, and G. Shen, "Use it free: Instantly knowing your phone attitude," in Proc. 20th annual Int. Conf. Mobile Computing and Networking, Maui, Hawaii, 2014, pp. 605–616.

[24] R. LiKamWa, B. Priyantha, M. Philipose, L. Zhong, and P. Bahl, "Energy characterization and optimization of image sensing toward continuous mobile vision," in Proc. 11th annual Int. Conf. Mobile Systems, Applications, and Services, Taipei, Taiwan, 2013, pp. 69–82.