Polarization-based Visible Light Positioning

Zhice Yang, Zeyu Wang, Jiansong Zhang, Chenyu Huang,
Qian Zhang, *Fellow, IEEE*

Abstract—Visible Light Positioning (VLP) provides a promising means to achieve indoor localization with sub-meter accuracy. We observe that the Visible Light Communication (VLC) methods in existing VLP systems rely on intensity-based modulation, and thus they require a high pulse rate to prevent flickering. However, the high pulse rate adds an unnecessary and heavy burden for receiving with cameras. To eliminate this burden, we propose the polarization-based modulation, which is flicker-free, to enable a low pulse rate VLC. In this way, we make VLP light-weight enough even for devices with a camera.

This paper presents the VLP system PIXEL, which realizes our idea. In PIXEL, we developed 1) a novel color-based modulation scheme to handle user's mobility and 2) a sensor-assisted positioning algorithm to ease user's burden in capturing multiple location anchors.

Our experiments based on the prototype show that PIXEL can provide 10 cm-level real-time VLP for wearables and smartphones with camera resolution as coarse as 60 pixel $\times$ 80 pixel and CPU frequency as low as 300 MHz.

Index Terms—Indoor Localization; Visible Light Communication; Polarization; Mobile Devices; Wearables

1 Introduction

We envision indoor positioning as being an indispensable feature of smartphones and wearables (*e.g.*, smart glasses) in the near future to support indoor navigation as well as plenty of location based services in shopping malls, supermarkets, office buildings, *etc.* To realize this goal, we need technologies that can provide high positioning accuracy as well as being light-weight enough to run in resource-constrained mobile devices, as these devices (especially wearables) are normally equipped with CPU, memory and camera that are optimized more for power-efficiency than performance.

Visible Light Positioning (VLP) [18], [19] is an emerging positioning technique that broadcasts anchor locations through *Visible Light Communication* (VLC). Benefited by densely deployed indoor lamps, VLP can easily achieve sub-meter accuracy in indoor localization. Compared with Wi-Fi and other RF based approaches which normally provide accuracy in meters, VLP holds the promise for beneficial applications such as retail navigation and shelf-level advertising in supermarkets and shopping malls.

However, we observe that in existing VLP systems, receiving devices are heavily burdened in VLC to avoid light flickering, which is caused by their intensity-based modulation method. As human eyes are sensitive to low rate changes in light intensity, the lamps have to transmit pulses at a high rate (over 1 kHz) to prevent flickering. Since the pulse rate far exceeds the camera's normal sampling capability (30fps), the design of the receiving side has to incorporate new hardware to mobile devices – additionally installs a customized light sensor, and relies on received signal strength techniques which are difficult to calibrate [19], [32]. Several recent approaches leverage the rolling shutter effect of the CMOS image sensor to decode high-rate pulses from the high-resolution image [18], [22], [26], [31], [33]. Although its effectiveness has been shown in high-end smartphones with high-resolution cameras, the decodable distance is very limited in middle-end smartphones with 10 megapixel cameras or smart glasses with 5 megapixel cameras. Moreover, the high-resolution image incurs huge computation which requires a long processing time or cloud off-loading, therefore it lengthens response time, increases energy cost and sacrifices reliability.

The heavy burden on receiving devices motivated us to try to eradicate the light flickering. Our idea is to modulate the light's polarization instead of its intensity for communication. This paper presents our VLP system PIXEL, which realizes this idea. PIXEL enables a light-weight VLP that is even affordable by wearables (Google glass), without incurring hardware modifications or computational off-loading. PIXEL also makes other types of illuminating light beyond

- *Z. Yang is with School of Information Science and Technology, ShanghaiTech University, China. Email: yangzhc@shanghaitech.edu.cn*
- *Z. Wang is with Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. Email: zwangas@connect.ust.hk*
- *J. Zhang is with Alibaba Group, China. Email: jiansongzhang@live.com*
- *C. Huang is with Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. Email: chuangak@connect.ust.hk*
- *Q. Zhang is with Department of Computer Science and Engineering, Hong Kong University of Science and Technology, Hong Kong. Email: qianzh@cse.ust.hk*

 DOI: 10.1109/TMC.2018.2838150

LED light (even sun light, §3) usable for transmitting location anchors, therefore eliminating the potential barriers to its deployment.

The design of PIXEL's polarization-based VLC is inspired by the *Liquid Crystal Display* (LCD). We borrow *polarizer* (§2.2.1) and *liquid crystal* (§2.2.2) from LCD as PIXEL's major components. A challenge in PIXEL's design is that the SNR in the channel of the polarization-based VLC differs dramatically in different receiving orientations. To address this problem, we propose to add a *dispersor* to PIXEL's VLC transmitter and employ a novel modulation scheme, called *Binary Color Shift Keying* (BCSK) (§3.1). With the help of these two modulation techniques, PIXEL is able to achieve reliable channel quality despite the user's mobility. In the positioning process, it is well known that accurate VLP must be performed with at least 3 location anchors, but our observation in real practice shows less anchors can be captured in general. PIXEL incorporates novel algorithm design to relieve user's burden in capturing enough VLP anchors to satisfy the positioning requirement. Specifically, PIXEL leverages accelerometer in client devices to provide additional positioning constraints and only 2 location anchors are required for accurate positioning (§3.2). Moreover, we formulate the accelerometer-assisted VLP positioning into a nonlinear least squares optimization problem and achieve orders of magnitude performance gain.

We implement PIXEL's VLC transmitter with commodity components which can be simply deployed by attaching to any light sources. We develop a software program for both Android smartphone and Google glass, they can receive VLC signals through a polarizing film in front of their cameras. We summarize the evaluation results and our contributions as below:

- We enable polarization-based VLC between PIXEL's transmitters and camera-equipped smart devices. Reliable communication over a long distance of 10-meters only requires an image/video capturing resolution of 120 pixel$\times 160$ pixel. As far as we know, we propose the first system design that leverages polarization-based VLC in open space.
- We enable a sensor-assisted VLP algorithm. It only requires 2 VLP anchors to perform accurate 3D positioning with 90% error less than 10cm. To the best of knowledge, our algorithm requires the fewest VLP anchors to achieve such accuracy level.
- We enable realtime visible light positioning in resource-constrained devices and the maximum positioning delay is less than 3 seconds. Even when the CPU frequency is as low as 300MHz, PIXEL's light-weight design works properly.

2 OVERVIEW

This section provides an overview of PIXEL. We first introduce the background of VLP and the flickering problem which motivates our work. Then we introduce the basics of polarization-based VLC, including techniques for generating/detecting polarized light and modulating light's polarization. Finally, we introduce the idea of PIXEL' system which is inspired by the design of one pixel in LCD screen.

2.1 VLP and the Flickering Problem

Visible Light Positioning (VLP) relies on *Visible Light Communication* (VLC) to broadcast location information through modulated light beams. The receivers, which are usually mobile devices carried by humans, use the light sensor [19] or the camera [18] to receive the location information as well as measuring their relative position to the lamps to perform fine-grained positioning. The mechanism is very similar to Wi-Fi positioning except that the visible light is used as the carrier to carry beacon messages instead of the microwave signal. Benefiting from the high density of the indoor lamps, visible light positioning can achieve much better (sub-meter level) accuracy than Wi-Fi positioning.

A major headache in VLP is the flickering problem. In current VLP designs, communication is based on modulating the light's intensity. Optical pulses are generated and data bits are encoded using pulse rate or pulse width. The varying light intensity causes flickering which can make people uncomfortable or even nauseous [13], [24]. In order to avoid the negative impact of flickering, a high pulse rate has to be used to make the pulses unnoticeable, though the message size required for localization is small. For example, the existing designs use over-1kHz pulse rates to transmit 7-bit beacons [18] or over-10kHz pulse rates to transmit 16-byte beacons [19].

The forced high pulse rate indeed adds a heavy burden to receiving devices. Considering the camera as the most common visible light sensor, the normal frame rate of 30 frames per second (fps) [16] is much lower than the pulse rate, which means it is far from sufficient to recover the transmitted pulses if we take one light intensity sample from every frame[1]. An interesting idea to tackle this problem is to leverage the rolling shutter effect of CMOS image sensor [18], [26]. The effect means rather than scanning a scene all at once, the image sensor scans one line at a time. Rolling shutter increases the sampling rate by up to thousands of times (from tens of fps to tens of thousands of lines per second), therefore making it possible to cover the pulse rate. However, according to our experience, the usefulness of rolling shutter is limited for two reasons. First, the camera must have very high resolution if it is not close enough

1. Even cameras specially designed for capturing slow-motion video, for example, the 240-fps camera in the latest iphone6, the frame rate is still not sufficient to cover the over-1kHz or higher pulse rate. We also do not expect this slow-motion capable camera to be widely equipped into low-end phones and wearables.

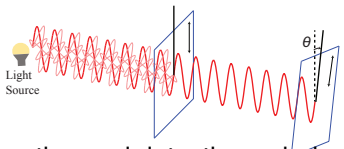


Fig. 1. Generating and detecting polarized light.

to the lamp. This is because the lamp is only a small portion of the image, but a certain number of occupied lines/pixels are required to carry a message. In other words, if the receiving device does not contain a high-resolution camera, the communication distance will be short. For example, on a Samsung Galaxy 4 with a 13-megapixel camera, we can reach up to 1.6 meters, and with Google glass's 5-megapixel camera, we can reach up to 0.9 meters. Second, processing high-resolution image requires very high computational power. As a consequence, the existing design [18] relies on a cloud/cloudlet server to process the image and decode the message. Besides camera and rolling shutter, another solution [19] leverages a customized light sensor which is capable of taking high-rate samples on light intensity. Different from camera-based solutions which use the angle-of-arrival based positioning algorithm, the light sensor based solution relies on *received signal strength* (RSS) technique to estimate the distance, this however are difficult to calibrate. Both hardware customization and RSS calibration are heavy burdens for VLP clients.

In this paper, we propose to exploit light polarization to realize low-pulse-rate VLC, so as to avoid the heavy burdens mentioned above and enable a light-weight VLP design that is affordable on wearable devices.

2.2 Basics of Polarization Based VLC

Polarization is a property of light waves that they can oscillate with more than one orientation. Similar to oscillating frequency which is another property of light, polarization preserves during light propagation in the air. However, different from oscillating frequency which determines color and thus can be perceived by human eyes, light's polarization is not perceptible [15]. This human-imperceptible nature has been proved very useful in that the polarization has led to the invention of Liquid Crystal Display (LCD) techniques, in which screens indeed always emit polarized light [30].

In the following subsections, we introduce two basic techniques in polarization-based VLC which solve the problems of generating/detecting polarized light and modulating light's polarization.

2.2.1 Generating/Detecting Polarized Light

Most of the common light sources, including sun light and man-made lighting systems, *e.g.* fluorescent light, incandescent light, *etc.*, produce unpolarized light which contains light of mixed polarization. The technique to convert unpolarized light into polarized light, called *polarizer*, is mature and very inexpensive. For example, the polarizer we use in this paper is a thin film, which costs roughly $0.01 per lamp. Polarizer is also widely used to detect polarized light and filter out certain unwanted light[2]. The theory behind polarizer is the Malus' law. That is, if the polarized light incident to a polarizer has intensity I_0, the passed light will have intensity I_θ, which is determined by the bearing angle θ between the polarization of the light and the polarizing direction of the polarizer:

$$I_\theta = I_0 \cos^2 \theta \tag{1}$$

Therefore, if the incident light has polarization in a direction parallel to the polarizer, *i.e.* $\theta = 0$, the light will pass through the polarizer without being attenuated. Otherwise, if their directions are perpendicular, *i.e.* $\theta = 90°$, the incident light will be entirely blocked by the polarizer. Malus' law can be applied to the case of converting unpolarized light into polarized light by considering unpolarized light to be a mix of polarized light with all possible polarizations. Figure 1 illustrates the process of generating and detecting polarized light according to Malus' law.

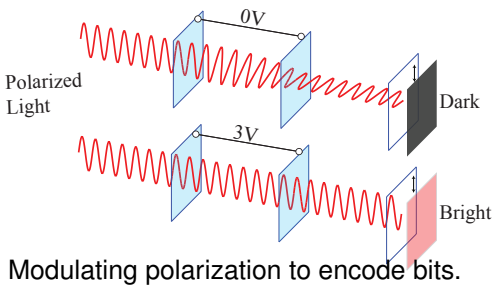


Fig. 2. Modulating polarization to encode bits.

2.2.2 Modulating light's Polarization

Modulating polarization of visible light using *liquid crystal* (LC) is also a mature and inexpensive technique. In this paper, we use Twisted Nematic (TN) liquid crystal which is very popular in the LCD screens of personal computers, digital watches, *etc.* TN liquid crystal costs only $0.03 per cm^2 or $1.5 per lamp in our design.

We explain how to modulate polarization to encode bits using Figure 2. The two surfaces, between which a variable voltage is applied, are the surfaces of a TN liquid crystal layer. TN liquid crystal has the characteristic that when a voltage is applied, its molecules will be realigned – twisted/untwisted at varying degrees depending on the voltage. Most commodity TN liquid crystal components are manufactured to have a 90° twist when no voltage is applied. Therefore a beam of polarized light will be rotated by 90° while passing through the liquid crystal, as shown in the upper sub-figure. If we use a polarizer parallel to the

2. For example, both polarized sunglasses and photographic polarizing filter are used to eliminate unwanted reflections.

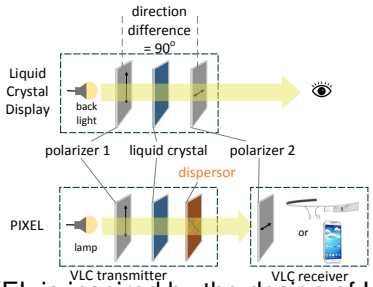


Fig. 3. PIXEL is inspired by the design of LCD.

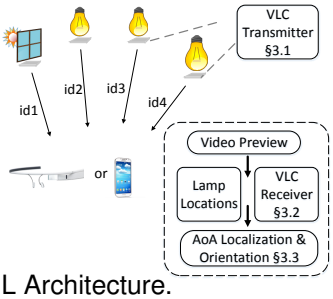


Fig. 4. PIXEL Architecture.

incident light to detect the light, as shown in the figure, the light will be entirely blocked by the polarizer. Then when we increase the voltage, the liquid crystal will untwist and the polarizer will become brighter. Once a certain voltage level is reached, *e.g.* $3V$ as in the figure, the liquid crystal will become totally untwisted and the polarizer will show the brightest state. In this way, we can tune the applied voltage to modulate the polarization and encode '0' and '1' using dark and bright states[3].

2.3 PIXEL – A Pixel of LCD

Our system, PIXEL, is inspired by the design of the pixels in LCD. As shown in the upper sub-figure of Figure 3, a pixel of LCD consists of two polarizer layers and a liquid crystal layer in between. The backlight generates unpolarized light which is converted into polarized light by polarizer 1. The liquid crystal layer twists/untwists the light's polarization according to the applied voltage. Polarizer 2 has a fixed angle of 90° in the polarizing direction with polarizer 1. Therefore, the strength of the emitted light is controlled solely by the voltage applied on the liquid crystal layer. Finally, a lot of pixels each of which has an independent control voltage constitute an LCD screen.

The lower sub-figure illustrates the VLC design of PIXEL. We convert a lamp into PIXEL's VLC transmitter by adding a polarizer and a liquid crystal layer. The transmitter thus emits polarized light. Light's polarization can be changed by the voltage applied to the liquid crystal layer. The change of light's polarization cannot be perceived by human eyes or directly captured by cameras. PIXEL's VLC receiver is a mobile device which equips a camera, *e.g.* smart glasses and smartphones. We add another polarizer to the camera for the purpose of detecting the changes of light polarization. As discussed in §2.2.2, the polarizer converts polarization changes into brightness changes that can be captured by the camera. We notice that the effect of adding polarizer to the camera, considering normal photo taking and video recording, is very similar to the use of polarizing filter in photography [21], *i.e.*, the polarizer does not bring noticeable changes except a slightly reduction in the strength of the ambient light.

However, from Figure 3, we can observe an important difference between PIXEL and LCD. In PIXEL, the polarizing direction between polarizer 1 and polarizer 2 does not always remain constant, because the receiving device may move and rotate with the person who holds it. This varying angle is one of the major challenges in our design. In the worst case the receiving camera will capture almost the same light intensity in the two polarization states. To tackle this problem, we leverage the camera's ability of capturing chromatic images and add a dispersor to disperse light of different colors to different polarizations. The details as well as other challenges and solutions are elaborated on in the next section.

3 PIXEL Design

PIXEL presents a system design which makes visible-light based accurate indoor positioning affordable in wearables and low-end smartphones. As shown in Figure 4, every location beacon periodically sends out its identity through visible light communication. A wearable device or smartphone captures multiple beacons using the camera's video preview. Relative locations of beacons are obtained directly from the video. The identities of beacons are decoded by PIXEL's VLC receiver. Finally, with the help of a location database indexed by beacon identities, both location and orientation of the receiving device are determined using *angle-of-arrival* (AoA) based algorithms.

In PIXEL, it is possible to make use of any type of visible light source to construct a location beacon. As illustrated in Figure 4, a location beacon may consist of a lamp (not just the LED lamp, but any lamp) and a PIXEL's VLC transmitter. A location beacon may also consist of a window with incident sun light and a PIXEL's VLC transmitter.

In the following of this section, we elaborate on three main components of PIXEL, *i.e.* VLC transmitter, VLC receiver and the AoA based localization/orientation algorithm, to explain the challenges as well as our solutions in the design.

3.1 VLC Design

The VLC design in PIXEL is borrowed from LCD, as has been described in §2.3. The VLC transmitter

3. In theory, we can encode more bits using multiple levels of brightness. However, in this paper we only encode 1 bit to get rid of complicated calibration.

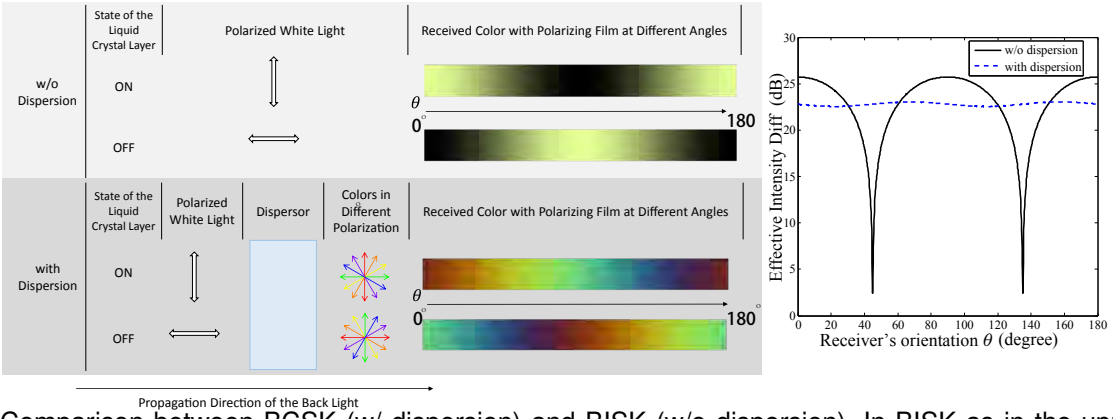


Fig. 5. Comparison between BCSK (w/ dispersion) and BISK (w/o dispersion). In BISK as in the upper sub-figure, the captured white light in receiving camera can only be quantified using light intensity. In BCSK as in the lower sub-figure, a dispersor is added in the VLC transmitter, the captured chromatic light can be quantified using a vector in RGB domain (*i.e.* color). The right sub-figure shows in BISK, intensity-diff of two encoded states largely fluctuates with receiver's orientation, whereas in BCSK, the effective intensity-diff is stable.

contains a polarizer and a liquid crystal layer that can convert the unpolarized illuminating light into polarized light and modulate the light's polarization using a control voltage. Then when the light propagates to receiving devices, polarization changes will be converted into intensity changes by another polarizer and identified by the receiving camera. Finally, the VLC transmitter also contains a simple electronic module to encode information, which translates the beacon's location identity into a series of control voltages.

A big challenge to PIXEL's VLC comes from the receiver's mobility, as has been briefly mentioned in §2.3. Unlike the two fixed polarizers in LCD, in PIXEL, though the polarizer in VLC transmitter is fixed (attached to the lamp), the polarizer in the receiver is mobile. Therefore, the two polarizers could end up with arbitrary angles in their polarizing directions. Since this angle determines the intensity of light, which is finally captured by the receiving camera, angle ambiguity will lead to a varying *intensity difference* (intensity-diff) between '1' and '0' states, meaning varying *signal-to-noise-ratio* (SNR) for communication. In the worst case, the intensity-diff will approach zero, therefore the encoded states will be totally undecidable. In order to quantify the relation between the intensity-diff and the polarizing angle between two polarizers, we build a mathematical model using Malus' law. Theoretical results based on this model are shown in the rightmost side of Figure 5. We can see when the angle is roughly in a 20° range around 45°(or 20% possibility), the intensity-diff drops quickly towards zero and can hardly fulfill the SNR requirement for communication. We also verified the result using experiments, in the angle of 45°, the intensity-diff is indeed very low.

A possible solution is to ask users to rotate the receiving devices. However, we believe it would be a cumbersome task. Since the polarizing angle is not perceivable by eyes, there is hardly any information to guide users on how to rotate. Also, rotating wearable devices could require big weird gestures.

To tackle this problem, we exploit the camera's capability of taking chromatic video. Recall the rainbow, the white light consists of multiple colors of light. The basic idea is to disperse the polarized light emitted from the liquid crystal layer into different colors, so that the receiving camera can capture different colors in different orientations after the filtering of the receiving polarizer. We fine tune the color dispersion to guarantee that, in any orientation, the captured colors in two encoded states will be largely different. We term this color-based modulation as *Binary Color Shift Keying* (BCSK). Next, we describe the dispersor and BCSK in details.

3.1.1 Dispersor and BCSK

Some materials have a property called *optical rotatory dispersion* (ORD), *i.e.* they can rotate the light's polarization differently with different frequencies (colors) of light [5]. These substances may be solid such as quartz, liquid such as alcohol or gas such as Hexadiene. This property has been studied for decades and widely used in industry, chemistry and optical mineralogy. For example, the method for measuring the blood sugar concentration in diabetic people is based on the ORD property of the sugar. In this paper, we use a liquid substance called Cholest [8] as our dispersor for the ease of fine-tuning the thickness and dispersion.

We use Figure 5 to illustrate BCSK by comparing it with *Binary Intensity Shift Keying* (BISK). We can see the beam of the polarized white light emitted from the liquid crystal is dispersed into multiple beams of polarized light with each having a unique color and polarization. The receiving polarizer passes through one beam of light according to its polarizing direction and filters out other beams. Therefore, only a certain color is captured by the receiving camera. When the

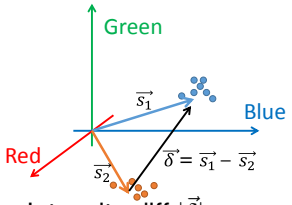


Fig. 6. Effective Intensity-diff $|\vec{\delta}|$.

encoded state is changed by the control voltage on the liquid crystal layer, the color of the captured light changes as well. Although the captured color still differs with the orientation of the receiving device, with proper dispersion we can guarantee the received colors largely differ between two encoded states in any orientation.

Next, we describe how to tune the dispersion for the purpose of guaranteeing a reliable SNR in all orientations.

3.1.2 Tuning Dispersion and Optimizing SNR

Dispersion tuning is performed by choosing the best thickness for disperor. The best thickness is mainly determined by the substance of dispersor. The spectrum of the illuminating light may slightly affect the best thickness, but this effect is very small because most illuminating light sources emit white light which represents a similar flat spectrum. When the substance of dispersor and the type of light is determined, the best thickness is also determined.

The target we optimize for is a metric called *effective intensity-diff*. It quantifies the color difference between two encoded states which is measured by camera. Therefore, it represents the received signal strength of the communication channel. Specifically, as shown in Figure 6, the colors captured by camera are mapped into a {*Red, Green, Blue*} (RGB) space. Given certain receiver orientation, the two colors in two encoded states can be indicated using two vectors $\vec{s_1}$ and $\vec{s_2}$. The difference between two colors is therefore $\vec{\delta} = \vec{s_1} - \vec{s_2}$. Then, effective intensity-diff is defined as $|\vec{\delta}|$.

We formulize the optimization problem and solve the best thickness L_{best} as follows.

When a beam of polarized light passes through a dispersor, it is dispersed into multiple beams of polarized light in which each has a unique frequency (color). The polarization of a beam with frequency f is rotated by an angle $\Delta\phi_f$, which is determined by:

$$\Delta\phi_f = n(f)L, \tag{2}$$

where L is the thickness of dispersor, $n(f)$ is a function of f which describes the rotation of light polarization in a unit thickness. $n(f)$ describes the property of a substance.

After the filtering by the receiving polarizer, the light captured by camera is described with a vector $\{I_R, I_G, I_B\}$ in RGB space. $\{I_R, I_G, I_B\}$ can be derived

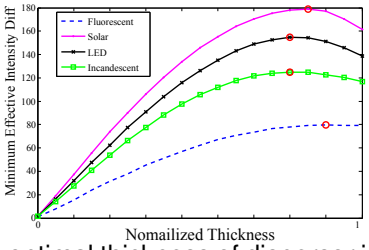


Fig. 7. The optimal thickness of dispersor is almost the same with various types of illuminating light, as most illuminating light is white light.

using the Malus' law:

$$I_N(\theta, L) = \int_{f_{min}}^{f_{max}} p_N(f)\cos^2(n(f)L - \theta)\, df, N \in \{R, G, B\},$$

where θ is the polarizing angle between transmitter's polarizer and receiver's polarizer. f_{min} and f_{max} are the upper/lower frequency bounds of the light's spectrum. $p_N(f)$ describes the light spectrum projected on the dimention N of RGB space. $p_N(f)$ is determined by the camera's chromatic sensor. From this formula, we can observe that the light captured by camera is determined by both optical rotatory dispersion ($n(f)L$) of the substance and the light spectrum ($p_N(f)$), as mentioned above.

The final goal of the optimization problem is to maximize the minimal effective intensity-diff in all receiving orientations. Considering the two encoded states between which the polarizing angle is rotated by 90°, we can formulize the optimization problem as follows:

$$\arg\max_{L}\min_{\theta}\sqrt{\sum_{N\in\{R,G,B\}}\left(I_N(\theta, L) - I_N(\theta + 90°, L)\right)^2}.$$

To solve this problem and get L_{best}, we need to know $n(f)$ and $p_N(f)$ in advance. $n(f)$ and $p_N(f)$ can be obtained from specification or calibration. In our work, we calibrate the $n(f)$ of our dispersor by ourselves and derive $p_N(f)$ from specification [3] and calibration results in literature [1]. We found $p_N(f)$ almost does not affect the vlaue of L_{best}, as shown in Figure 7[4]. We believe that it is because most illuminating light sources emit white light which has a similar flat spectrum.

To compare BCSK with BICK, we calculate the effective intensity-diff under L_{best} in all possible receiver orientation (represented by θ). Results are shown in the rightmost of Figure 5. We can see the effective intensity-diff is quite stable in all the receiver orientations.

To conclude, the dispersor in PIXEL's VLC transmitter and the color-based modulation scheme BCSK well

4. We use the normalized thickness to plot the trend for dispersors with different materials. Given the condition that the $n(f)$ is approximately linearly related to f, the trend of the Figure holds. Length "1" means the the difference of the rotated angles between f_{min} and f_{max} is 180°: $|\Delta_{f_{max}} - \Delta_{f_{min}}| = 180°$ and length "0" means no difference in rotation, *i.e.* L=0.

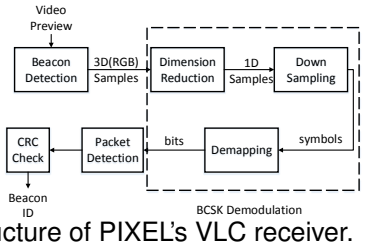


Fig. 8. Structure of PIXEL's VLC receiver.

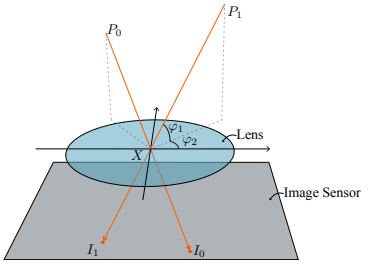


Fig. 9. Imaging geometry in camera-based AoA positioning. P_1 and P_2 denote VLC transmitters/location anchors. Light passing through simple biconvex lens keeps its direction unchanged. I_1 and I_2 are anchors' image projected on image sensor. Each location anchor adds two arriving angle constraints φ_1 and φ_2.

solve the problem of varying SNR caused by users' mobility.

3.1.3 VLC Receiver

PIXEL's VLC receiver decodes beacon's identity from camera's video preview. Figure 8 shows the structure of the VLC receiver.

The first step is beacon detection. In order to save computation, we first downsample the video to contain less pixels. According to our experience, 100×100 pixels will be sufficient for decoding in a distance within 10 meters. Then we calculate the difference between every two consecutive frames to detect the VLP beacons. In order to make the detection reliable, we add up the differences from 10 consecutive frames. Usually a detected beacon contains 10 or more pixels which contain its color value. We finally average these pixels to generate a 3-dimensional ($\{R, G, B\}$) color sample from every video frame. Once a beacon is detected, one frame without downsampling is buffered for calculating its image location for the localization algorithm in §3.2.

As illustrated in Figure 6, these 3D samples in $\{R, G, B\}$ space are mainly distributed around two clusters, which represent the two encoded states. Since there are only two clusters/states, for the ease of further processing, we perform dimension reduction to convert 3D samples into 1-dimensional samples as the first step of BCSK demodulation. The requirement of dimension reduction is to preserve the distance between samples of the two encoded states. Therefore, we perform reduction by projecting these 3D samples to the difference vector between the two states, *i.e.* $\vec{\delta}$ in Figure 6. The projection from a 3D sample $\vec{s}$ to 1D sample s is calculated by: $s = \vec{s} \cdot \vec{\delta}/|\vec{\delta}|$. To determine $\vec{\delta}$, we calculate difference vectors between every two consecutive 3D samples, then take the average of the difference vectors with large norm and right direction as $\vec{\delta}$.

In PIXEL's transmitter, we set the symbol rate roughly the half of the camera's sampling rate for the purpose of meeting the sampling theorem. As the low cost cameras in wearables are not stable in the sampling interval, we make use of the OS clock to timestamp each video frame and use that time to aid downsampling [29].

The output of the downsampling module is symbols. Since the symbols are in 1 dimension (scalar), demapping is equivalent to bit arbitration which determines '1' or '0' from every symbol. Similar to classic 1-bit arbitrator designs, we set the average of all symbols as the threshold.

The final step is to recover packets from the bit stream. In PIXEL, we use a 5-bit header for packet detection and synchronization. Since the packet is short, we use a 4-bit CRC for error detection and do not perform channel coding or forward error correction. More details about the packet design is shown in §4.

3.2 Positioning Method

PIXEL's clients leverages its camera for positioning. Since the direction of light beams passing through the center of simple biconvex lens keeps unchanged [18], the geometry relation (Figure 9) contains information about the arriving angles of transmitters. With enough transmitter anchors, the location of the camera can be determined through Angle of Arrival (AoA) methods.

Existing AoA methods [18], [19] require at least 3 transmitter anchors to achieve accurate positioning results. Even though light lamps are ubiquitous, our field measurement indicates less lamps can be captured in general. In most areas of our campus, 2 or less lamps can be captured by front cameras of smartphones with normal holding positions. Similar results are also observed in other office and shopping areas. There are mainly two reasons limit camera's coverage. First, cameras for mobile use are not designed for wide field of view (FoV). Second, normal corridors have limited height. As an example, the front camera of iPhone 6 can only cover $9m^2$ of the ceiling area $3m$ above of it. In order to perform accurate positioning, tilting or pointing device with weird postures may be frequently required to capture more than 3 anchors.

However, less anchors, says 2, bring ambiguities in locations in AoA methods. The location of the camera X is in 3D space and has 3 degrees of freedom (DoFs). Although each anchor has 2 constraints in angles φ_1 and φ_2, 2 anchors with 4 constraints are not enough to uniquely determine the location X with 3 DoFs. This is because the camera is a 3D object, its orientation (Pitch, Roll, Yaw) with 3 DoFs also affects the imaging

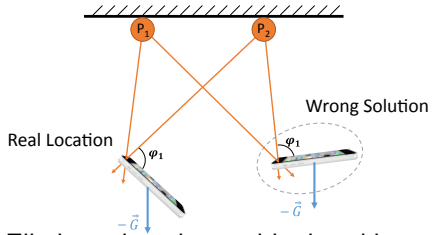


Fig. 10. Eliminate location ambiguity with gravity measurement. This example shows geometry constraints from 2 location anchors are not enough to determine accurate 3D location, because the wrong but feasible solutions have exactly the same arriving angles as the real location (the one out of four angles, φ_1 is depicted). However, the ambiguity can be eliminated by considering the direction of the gravity $-\vec{G}$.

area of the anchors. In other words, AoA constraints from anchors are partially related to the camera's orientation. Therefore, 2 anchors with 4 constraints are not enough to fully resolve 6 DoFs from the location and orientation of the camera. A simple example with 2 anchors in Figure. 10 evidents the above analysis.

3.2.1 Sensor Assisted AoA

PIXEL's design addresses the above practical problem by incorporating the information from sensors in the same smart devices to add constraints for positioning. We choose the wildly-available sensor: accelerometer to measure the direction of the gravity, so that the tilting angle between the camera and the gravity can be used to eliminate wrong locations as Figure 10 shows. With this restriction, DoFs of the orientation are reduced by 2, so that the orientation and location of the camera only contain 4 DoFs in total. Therefore, 2 anchors with 4 constraints are enough for PIXEL to perform accurate positioning. In the followings, we formulate the problem and describe our algorithm in detail.

We use the notations as Figure.9. X is the location of the center of the camera lens, and we also treated X as the location of the client. The transmitter anchors are located at P_k with corresponding image points I_k on the image sensor. As the light beam is a straight line, the direction of $\overrightarrow{P_kX}$ and $\overrightarrow{XI_k}$ must be the same:

$$\frac{\overrightarrow{P_kX}}{|\overrightarrow{P_kX}|} = \frac{\overrightarrow{XI_k}}{|\overrightarrow{XI_k}|}. \tag{3}$$

Equation (3) is the relation between anchors' locations and their image points. In this equation, we emphasize that the left part can be expressed in the Earth coordinates: $\overrightarrow{P_kX}\big|_e = [x_1, x_2, x_3]^T - [p_{k1}, p_{k2}, p_{k3}]^T$, and the right part can be calculated in the local coordinates of the camera: $\overrightarrow{XI_k}\big|_l = [i'_{k1}, i'_{k2}, i'_{k3}]^T - [x'_1, x'_2, x'_3]^T$. Moreover, the two parts are related by the coordinate transformation matrix R^l_e. Without loss of generality we take $[x'_1, x'_2, x'_3]^T = [0, 0, 0]^T$ and equation (3) can then be written as:

$$\left| \frac{R^l_e \cdot [x_1 - p_{k1}, x_2 - p_{k2}, x_3 - p_{k3}]^T_e}{\sqrt{(x_1-p_{k1})^2+(x_2-p_{k2})^2+(x_3-p_{k3})^2}} - \frac{[i'_{k1}, i'_{k2}, i'_{k3}]^T_l}{\sqrt{i'^2_{k1}+i'^2_{k2}+i'^2_{k3}}} \right| = f_k(x_1, x_2, x_3, R^l_e) = 0. \tag{4}$$

Among these variables, $[p_{k1}, p_{k2}, p_{k3}]^T$ are informed by VLC beacons and $[i'_{k1}, i'_{k2}, i'_{k3}]^T$ can be calculated from the captured image by considering camera's parameters. Camera's location in the Earth $[x_1, x_2, x_3]^T$ is what we want to obtain but R^l_e is unknown.

To solve $[x_1, x_2, x_3]^T$ from equation (4), we derive the explicit expression of matrix R^l_e by considering the direction of the gravity. Since $R^l_e \cdot [1, 0, 0]^T = [n'_1, n'_2, n'_3]^T = \vec{N}\big|_l$, the first column of R^l_e is the North direction in local coordinates. Similarly, as $R^l_e \cdot [0, 0, 1]^T = [g'_1, g'_2, g'_3]^T = \vec{G}\big|_l$, the third column of R^l_e is the negative direction of the gravity. For the second column, $R^l_e \cdot [0, 1, 0]^T = R^l_e \cdot ([0, 0, 1]^T \times [1, 0, 0]^T) = \vec{G}\big|_l \times \vec{N}\big|_l$. Therefore, R^l_e can be written as:

$$R^l_e = \begin{bmatrix} n'_1 & g'_2 n'_3 - g'_3 n'_2 & g'_1 \\ n'_2 & g'_3 n'_1 - g'_1 n'_3 & g'_2 \\ n'_3 & g'_1 n'_2 - g'_2 n'_1 & g'_3 \end{bmatrix}. \tag{5}$$

In the above matrix, $\vec{G}\big|_l = [g'_1, g'_2, g'_3]^T$ can be measured by the accelerometer. As $\vec{N}$[5] is vertical to $\vec{G}$, it can always be represented in the normal plane of $\vec{G}$:

$$\vec{N}(\theta) = \vec{G}_{\perp 1} \sin\theta + \vec{G}_{\perp 2} \cos\theta. \tag{6}$$

$\vec{G}_{\perp 1}$ and $\vec{G}_{\perp 2}$ are two base vectors in the normal plane of $\vec{G}$, and they satisfy $\vec{G}_{\perp i} \cdot \vec{G} = 0$, $\vec{G}_{\perp i} \cdot \vec{G}_{\perp j} = 0$ and $\vec{G}_{\perp i} \cdot \vec{G}_{\perp i} = 1$. Since $\vec{G}$ is known, $\vec{G}_{\perp 1}$ and $\vec{G}_{\perp 2}$ can always be calculated.[6] θ is an unknown number and it describes the remaining one DoF in the camera's orientation. By combining equation (5) and (6), the transformation matrix has only one unknown parameter: $R^l_e = R^l_e(\theta)$. Then, equation (4) can be rewritten as: $f_k(x_1, x_2, x_3, \theta) = 0$.

With multiple input of P_k, I_k, we can obtain x_1, x_2, x_3 and θ when the sum in (7) achieves the minimum value:

$$\min_{x_1, x_2, x_3, \theta} \sum_k f^2_k(x_1, x_2, x_3, \theta). \tag{7}$$

Then the location of the PIXEL's client is just x_1, x_2, x_3 and the orientation of the client can be obtained by substituting θ in to (6).

So far, we have formulated the positioning problem into a typical nonlinear least square optimization problem. PIXEL adopts trust region algorithm with boundaries [10] to optimize the searching efficiency and reduce the computational overhead.

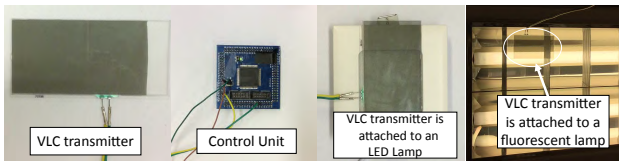


Fig. 11. Photos of PIXEL's components. From left to right: (1) VLC transmitter including a polarizing film, a liquid crystal layer and a dispersor. (2) The control unit with Altera Cyclon 2 FPGA. (3) VLC transmitter attached to a LED lamp. (4) VLC transmitter attached to a florescent lamp.

4 Implementation

Our implementation includes hardware part and software part.

Hardware. One VLC transmitter consists of three transparent optical layers and one control unit. The first layer is a linear polarized film, which functions as a polarizer to polarize the illuminating light. The second layer consists of two pieces of glasses with liquid crystal in between. We add electrodes on both glasses to control the applied voltage. This layer is a kind of simplified LCD with only one pixel, and many LCD manufacturers have the service for costuming LCD. The third layer is a dispersor. We prototype this layer by filling the optical rotation material (Cholest) into a glass box. We choose the liquid material because it is easy to tune the thickness, but solid materials with appropriate thickness could be a good choice for the real deployment. We adopt Altera's Cyclone II FPGA to control the voltage of the liquid crystal layer. The two electrodes of the liquid crystal layer are directly connected to two output ports of the FPGA to modulate the light's polarization.

Each PIXEL's transmitter works independently by facing to normal light sources such as solar light, LED light, fluorescent light and so on. As Figure 11 shows, normal light sources can thus be easily upgraded to have VLC ability with PIXEL's transmitter design.

The hardware of PIXEL's client devices requires no modification or customized accessories. The only requirement is a small piece of off-the-shelf linear polarizing film, and it is placed in front of the camera to detect and receive VLC signals.

Software. The packet format of VLC is designed for the indoor localization purpose. PIXEL's transmitter continuously broadcasts its location identity to receivers. The beacon packet contains a 5-bit preamble, an 8-bit data payload and a 4-bit CRC.

We implement the VLC decoding and the positioning algorithm in both Android system and an offline simulator[7]. We process the image data from the camera preview to decode VLC signal in realtime. The required parameters (*i.e.* the focal length of the lens and the size of the image sensor) for measuring $I_k|_l$ can either be read from the EXIF stamp of the camera or obtained from online specifications. Since our current implementation is based on Java, the processing time can be further reduced with C native implementation in Android.

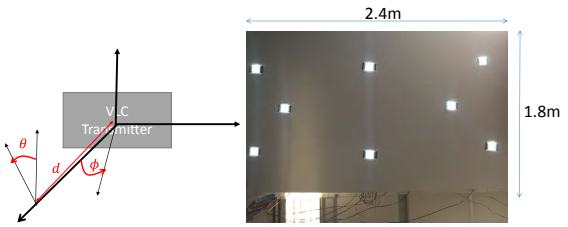


Fig. 12. Evaluation Setup. (a) is the settings for single VLC pair. d is the distance from the receiver to the transmitter. θ is the orientation of the camera. ϕ is the viewing angle of the transmitter. (b) is the testbed for the VLP evaluation.

5 Evaluation

In this section, we evaluate PIXEL's design based our prototype implementation. We first describe the evaluation setup, then we conduct experiments to verify the component design in PIXEL. After that, we thoroughly evaluate different impact factors on PIXEL's VLC subsystem and the computational cost of each part of the client application. Lastly, we present the overall performance of our VLP system.

5.1 Evaluation Setup

Evaluation involving one single pair of transmitter and receiver is performed in controlled conditions. Figure 12 (a) shows controlled parameters related to user's mobility. d is the distance between the transmitter and the receiver. θ is the receiver's orientation whose rotation plane is parallel to the receiver's plane. ϕ is the viewing angle whose rotation plane is vertical to the transmitter's plane. Other impact parameters such as the downsampled resolution[8] and the exposure time are also controlled in the evaluation. If not mentioned, we take $d = 3m$, $\theta = 0°$, $\phi = 0°$, downsampled resolution = 120×160 and exposure time = $1/100s$ as default settings.

Evaluation involving multiple pairs of transmitter and receiver is performed with the testbed in Figure 12 (b). We fix eight PIXEL's transmitters into a board which is elevated to the ceiling. The location of each transmitter is measured and mapped to the location ID in client's database. When the testbed is used for positioning evaluation, transmitters transmit their

5. $\vec{N}|_l$ can be measured from the magnetometer, but it contains large error in indoor environment.

6. We can first choose one random unit vector in the normal plane of $\vec{G}$ as $\vec{G}_{\perp 1}$, then $\vec{G}_{\perp 2} = \vec{G} \times \vec{G}_{\perp 1}$.

7. We extract Presentation Time Stamp (PTS) from `ffmpeg`(www.ffmpeg.org) to simulate timestamps of frames in recoded videos.

8. The downsampling is used to reduce processing overhead as we mentioned §3.1.3, we fix the downsampled resolution to certain values, such as 120×160. The reason is to normalize different resolutions across cameras.

location ID through frame defined in §4. Under the testbed, we grid another board to serve as ground truth reference for client's location.

We use controlled random bit stream for transmission to measure BER and SNR. In these experiments, the transmitter repeatedly transmits a random pattern with a length of 100 bits, while the receiver keeps logging the received color values and the decoded bits. Since the length of the random pattern is long enough, we determine the start of the transmitted pattern through correlation. Therefore, the decoded bits can be used to calculate BER by comparing with the known random pattern. With the known random pattern as ground truth, the received color values can be used to calculate SNR.

All the experiments are done without a fixed ambient light source. In the morning, the ambient light is solar light and becomes fluorescent light in the evening. We adopt the 5W 10cm$\times$10cm LED lamp as the light source of the transmitter for convenient setup. We also have performed similar evaluation experiments for fluorescent light and solar light, but we found little difference from LED.

5.2 Visible Light Communication

Experiment: We study PIXEL's VLC subsystem against five impacting factors caused by user mobility and device diversity: the distance d, the receiver's orientation θ, the viewing angle ϕ, the downsampled resolution and the exposure time of the camera. The experiment for certain factor varies that factor to all the reasonable values, and fixes other factors.

Results: Figure 13 (a) shows the SNR against the distance and the downsampled resolution. The downsampled resolution and the distance determine the number of pixels of the transmitter in the received image, and thus affect the SNR in a similar way. Therefore, we list them in the same figure for better illustration. From the results, our system achieves the SNR higher than 15dB under 120×160 resolution within $13m$, which is sufficient for most indoor localization scenarios. In addition to that, the VLC system works even when the downsampled resolution is as low as 30×40. Notice that the SNR of the downsampled resolution of 60×80 and 30×40 drops to 0 at $13m$ and $7m$ respectively. The reason is that our decoding algorithm needs enough pixels to extract color. However, when the distance from the transmitter increases, the transmitter's pixels in the image decrease. This is not a problem since even $13m$ is enough for VLP in a normal building. Moreover, the size of the transmitter or the intensity of the background light can easily be increased to satisfy the possible requirement.

Figure 13(b) shows the VLC performance under different orientations. Without the dispersor, the SNR becomes undecodable when the view angle between the transmitter and receiver is around 40° or 130°. This result is in accord with our previous analysis about effective intensity diff in §3.1. On the other hand, the result is improved a lot when the dispersor is deployed. The SNR of the received data is around 16dB in all the orientations, implying that the VLC quality is stable despite different orientations.

Figure 13(c) implies that the performance of VLC is stable under different viewing angles from -60° to 60°, which is sufficient for normal localization scenarios. From the results, we also observe that the variance of the SNR in the positive and negative sides are asymmetric. This is caused by the viewing angle limitation of the prototyping TN LCD, and can be solved by using better TN LCD or other LCD such as In-Plane Switching LCD.

Figure 13(d) shows the SNR against different exposure time settings of the camera. The result implies that there is no obvious relationship between exposure time and SNR. The reason is that our decoding algorithm will use different areas for decoding according to the exposure time. When under short exposure time, the amount of the received light from the transmitter is small and the decoding algorithm uses the center part of transmitter to extract color. When under long exposure time, the center of the transmitter is overexposed and displays white. However, the color can be extracted from the glow of the transmitter's image. Therefore, our VLC is robust under different exposure settings.

Figure 13(e) summarizes the data from all the experiments and get the relationship between BER and SNR. The theoretical curve for BCSK is calculated according to BPSK. However, since the camera of our smartphone is not stable enough and may occasionally drop frames during the capturing, the bit error rate is slightly higher than the theoretical result. Another reason is the number of samples for statistics is not enough. It can be observed from the two "horizontal lines" in the bottom of the Figure 13(e). Since each logged data only contains hundreds of bits on average, the upper "line" is plotted by cases with 2 error bits and the other "line" is formed by cases with 1 error bit.

Figure 13(f) shows the VLC throughput versus the number of transmitters. The rate is counted by the received payload, which does not contain the preamble and CRC. The result shows that the throughput increases linearly with the number of transmitters. It implies that concurrent transmissions don't affect the performance of our VLC. This is reasonable, since VLC transmitters can be differentiated through their locations in the captured image. Specifically, as we described in §3.1.3 , the receiver first determines the area of a VLC transmitter in the image. Then, the information of the VLC transmitter will only be extracted from chromatic values of that area. Since light signals are very directional, no obvious interference is

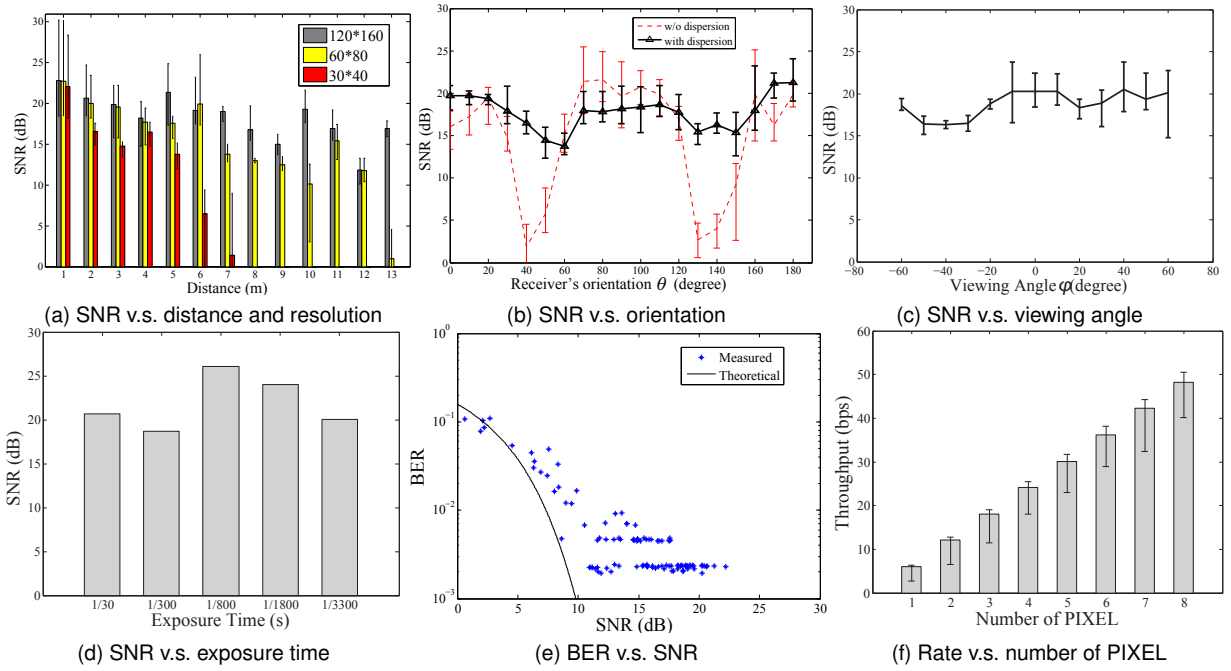


Fig. 13. PIXEL communication evaluation

observed when multiple transmitters are captured at the same time.

5.3 Visible Light Based Positioning

5.3.1 Computational Cost

Experiment: We study the computational cost of PIXEL's client application in detail. The evaluation is performed with Samsung Galaxy SII. Its CPU has dual cores with maximum frequency of 1200 MHz. We fix the frequency to 1200 MHz though an app `SetCPU`. It has two cameras and we use the back one with 8 Megapixel for this evaluation. The smartphone is fixed in one location to perform VLP and its distance to the testbed in Figure 12(b) is $3m$. We log the start and end time exclude the sleep period in each part of our program to calculate the processing time. This method may be affected by the CPU scheduling, so we close all the other applications and unnecessary services in the smartphone to keep the results as accurate as possible.

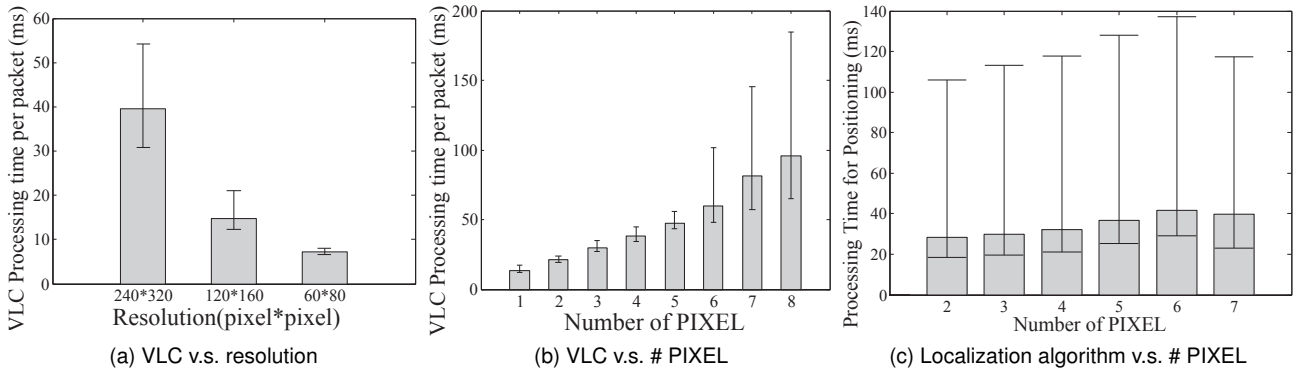
Results: Figure 14(a) shows the computational cost under different resolutions with single transmitter. Each packet is 17 bits and takes $1200ms$ air time for transmitting, but our decoding algorithm costs less than $50ms$ even when the resolution is 240×320. The result indicates our decoding algorithm performs fast enough. The figure also shows that it takes less time for processing when the downsampled resolution is lower. This is intuitive, because it takes less pixels for processing color extraction.

Figure 14(b) shows the time cost when the client is receiving from multiple transmitters. It shows that the total time of decoding 8 concurrent packets is less than $200ms$, which is still far less than the packet air time $1200ms$. The trend shows the decoding time increases linearly with the number of transmitters. This is because the decoding process for each transmitter is independent and scalable. We also note that the 8 cases share a small portion of basic processing cost, which is actually contributed by the shared downsampling process.

Figure 14(c) shows the time cost of the localization algorithm. Since it requires at least 2 anchors, the number of transmitters starts from 2. The result demonstrates that the localization can be finished within $150ms$ after the client receives enough beacons. The processing time slightly increases with the number of anchors, this is caused by the increased dimension in the object function.

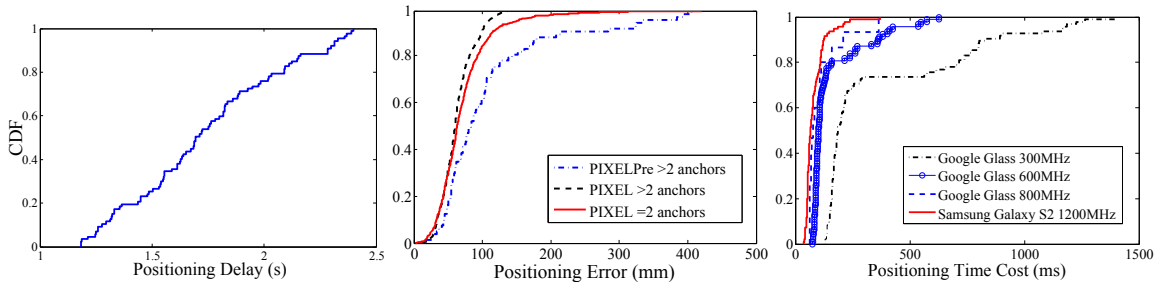
5.3.2 VLP Performance

Experiment: We study the overall VLP performance of PIXEL in three aspects: positioning delay, positioning error and positioning overhead. We use Samsung Galaxy SII and Google Glass as client devices while the transmitters in Figure 12(b) serve as VLP anchors. We place client devices to 100 random locations with arbitrary orientations and ensure that at least 2 transmitters can be captured by the camera. The start time and the end time of each test are logged for delay analysis. The localization results and the corresponding ground truth are recorded to estimate localization error. Like previous subsection, we log the processing time of the application to analyze the processing overhead. Since Google Glass has different frequency settings, we traversal all the configurations in our tests.



(a) VLC v.s. resolution (b) VLC v.s. # PIXEL (c) Localization algorithm v.s. # PIXEL

Fig. 14. Computational cost of each component



(a) Positioning Delay (b) Positioning Error (c) Processing time of VLP

Fig. 15. VLP overall performance

Results: Figure 15(a) shows that the average time from the launch of the application to a successful positioning is about $1.8s$. It is reasonable, because the time from the launch of the application to the earliest arrival of the beacon packet is uniformly distributed between 0 and 1200ms.

In Figure 15(b), we compare our positioning algorithm with existing approaches. Curve labeled as "PIXELPre" is the results from our conference version [29], which implements the AoA algorithm in [18]. We use Levenberg-Marquardt algorithm to optimize the execution [29], so that it can directly run on wearables. As it requires at least 3 anchors, the CDF only counts cases containing more than 2 anchors. It shows the localization error is less than 30cm in 90% test cases, which is in accord with the original results [18]. For the same test cases, our method achieves better accuracy even with only 2 anchors. In 90% of test cases, the positioning error is less than 10 cm. As expected, when more anchors are captured, the positioning accuracy can be further improved.

Figure 15(c) shows the CDF of the computational cost under different CPU settings. As mentioned above, the computational cost includes the VLC decoding and the localization algorithm. For normal mobile phones, all the tasks can be finished within 500ms. Even with a limited processing ability such as 300MHz in wearables, our design can finish all the calculation within 1000ms in 90% cases. Therefore, results validate that PIXEL can provide smart devices with light-weight and accurate localization ability.

6 Related Work

Visible Light Communication. VLC has been studied with different context and design goals. Comparing with radio communication using microwave or mmwave, VLC uses visible light which has much higher frequency and much wider bandwidth. Typical VLC research aims to provide very high link throughput communication [17] for the purpose of wireless access [23]. Normally the receiver requires a photodiode to receive the modulated light. Different from this goal, our design focuses on providing a light-weight VLC solution for camera-equipped and resource-constrained mobile devices.

In recent literatures [11], [18], [22], [26] it has been proposed to leverage rolling shutter effect of CMOS camera to conduct VLC. It makes use of the fact that LED light with different pulse rates can generate different line patterns in the images captured by camera. However, according to our experience, the rolling shutter based VLC requires high camera resolution and high computational power. In contrast, the VLC in PIXEL allows very low camera resolution and only requires light computation.

Using camera to decode static or dynamic barcode [14], [16], [27] is a special case of VLC. It has enabled a lot of interesting applications. Comparing to bar-code, PIXEL's VLC transmitter allows longer communication distance and is human imperceptible.

Visible Light Based Indoor Positioning. Indoor positioning has been studied for decades [4], [28]. Recent literature [18], [19], [22], [31], [32], [33] shows

that receiving location beacons from lighting infrastructure provides a promising means for accurate indoor localization. Comparing to existing designs, PIXEL aims for a more ambitious goal of enabling light-weight visible light positioning that is even afforable in wearables. Comparing to existing LED-based VLC transmitter, PIXEL leverages light polarization to avoid the flickering problem, therefore significantly reduces the requirement on transmitting pulse rate. As a result, the requirement on the VLC receiver is largely reduced that only low resolution camera is needed.

Polarization Based Modulation. VLC has been widely studied in the context of fiber communication and space communication. In the large body of the literature, some [6], [12], [25] has mentioned polarization based modulation. However, most of them do not contain a system design. Moreover, as far as we know, none of them has addressed the problem caused by the varying angle between transmitter and receiver because usually a fixed angle is assumed.

7 Discussion and Limitations

In this section, we discuss some limitations of our current system and potential directions for future work.

Bit Rate. In our current implementation, the baud rate of the VLC transmitter is set to 14Hz to allow most camera-equipped devices reliably decode. As a result, the bit rate of our implementation is 14bps. Though it is already sufficient for positioning purpose, the bit rate still can be increased in two ways. First, since the VLC transmitter is not the bottleneck (off-the-shelf LCD screens can achieve 1000Hz [2] refreshing rate), increasing the sampling rate of camera will make higher baud rate possible. For example, if Apple iPhone 6 that contains a 240-fps slow-motion camera is used as the receiver, the baud rate can be increased to 120Hz. Second, if we define more states of the polarizing direction, *i.e.* using higher modulation density, we can encode multiple bits into one symbol thereby increasing the bit rate.

Beacon Message Size. The beacon message size in our current design is 8-bit. Therefore, we can support up to 256 unique IDs. The message size is similar to existing design [18] and sufficient to cover one floor of normal buildings, *e.g.* shopping malls. In certain cases that require more/less unique IDs, the message size can be increased/decreased. As a consequence, the positioning delay will increase/decrease accordingly.

Power Consumption of Camera. Our VLC receiver relies on camera's video preview, therefore, we need to keep the camera on during the positioning process. A normal camera consumes around 300mW [20] power, which is comparable to the screen [7], [9]. However, since the overall positioning process is less than three seconds, the consumed power by the camera is limited.

Limitations. Since the polarization changes of the VLC transmission can be observed through polarizing films, people wearing polarized sunglasses can observe the light changes in VLC transmitters, which may cause interference. However, since wearing polarized sunglasses (normally for driving) in indoor is very uncommon, this is not a big issue.

Another concern is that the polarizing film in front of device's camera will reduce the intensity of the received light. The reduced light intensity under normal circumstance can be auto-compensated by dynamic sensitivity (iso) and exposure time of the camera. But in dark conditions, the polarizing film may reduce the imaging quality. Although the impact is limited, users can choose to paste polarizing film only when they need indoor localization service. As the service provider, administrators of supermarket, underground parking lot etc. can provide the film at the entrance.

Acknowledgement

This work was supported in part by the RGC under Contract CERG 16212714, 16203215, Contract ITS/143/16FP-A, and in part by the Guangdong Natural Science Foundation No. 2017A030312008. This work is supported in part by the ShanghaiTech Startup Fund F-0203-17-010, in part by the Shanghai Sailing Program 18YF1416700, and in part by Chenguang Program 17CG66 supported by Shanghai Education Development Foundation and Shanghai Municipal Education Commission.

References

[1] Colour rendering of spectra. http://www.fourmilab.ch/documents/specrend/, Mar 2015.

[2] Pi cell. http://www.liquidcrystaltechnologies.com/tech_support/Pi_Cell.htm, Mar 2015.

[3] Spectrum data. https://plot.ly/~arducordermini, Mar 2015.

[4] P. Bahl and V. N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000.

[5] N. Berova, K. Nakanishi, and R. Woody. *Circular dichroism: principles and applications*. John Wiley & Sons, 2000.

[6] R. Calvani, R. Caponi, and F. Cisternino. Polarisation phase-shift keying: a coherent transmission technique with differential heterodyne detection. *Electronics Letters*, 24(10):642–643, 1988.

[7] A. Carroll and G. Heiser. An analysis of power consumption in a smartphone. In *USENIX annual technical conference*, pages 1–14, 2010.

[8] S. Chandrasekhar. Optical rotatory dispersion of crystals. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 259(1299):531–553, 1961.

[9] X. Chen, Y. Chen, Z. Ma, and F. C. Fernandes. How is energy consumed in smartphone display applications? In *Proceedings of the 14th Workshop on Mobile Computing Systems and Applications*, page 3. ACM, 2013.

[10] T. F. Coleman and Y. Li. An interior trust region approach for nonlinear minimization subject to bounds. *SIAM Journal on optimization*, 6(2):418–445, 1996.

[11] C. Danakis, M. Afgani, G. Povey, I. Underwood, and H. Haas. Using a cmos camera sensor for visible light communication. In *Globecom Workshops (GC Wkshps), 2012 IEEE*, pages 1244–1248. IEEE, 2012.

[12] E. Dietrich, B. Enning, R. Gross, and H. Knupke. Heterodyne transmission of a 560 mbit/s optical signal by means of polarisation shift keying. *Electronics Letters*, 23(8):421–422, 1987.

[13] D. Giustiniano, N. Tippenhauer, and S. Mangold. Low-complexity visible light networking with led-to-led communication. In *Wireless Days (WD), 2012 IFIP*, pages 1–8, Nov 2012.

[14] T. Hao, R. Zhou, and G. Xing. Cobra: color barcode streaming for smartphone systems. In *Proceedings of the 10th international conference on Mobile systems, applications, and services*, pages 85–98. ACM, 2012.

[15] G. Horváth and D. Varjú. *Polarized light in animal vision: polarization patterns in nature*. Springer, 2004.

[16] W. Hu, H. Gu, and Q. Pu. Lightsync: Unsynchronized visual communication over screen-camera links. MobiCom '13, pages 15–26, New York, NY, USA, 2013. ACM.

[17] T. Komine and M. Nakagawa. Fundamental analysis for visible-light communication system using led lights. *Consumer Electronics, IEEE Transactions on*, 50(1):100–107, 2004.

[18] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta. Luxapose: Indoor positioning with mobile phones and visible light. MobiCom '14, pages 447–458, New York, NY, USA, 2014. ACM.

[19] L. Li, P. Hu, C. Peng, G. Shen, and F. Zhao. Epsilon: A visible light based positioning system. In *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*, pages 331–343, Seattle, WA, Apr. 2014. USENIX Association.

[20] R. LiKamWa, B. Priyantha, M. Philipose, L. Zhong, and P. Bahl. Energy characterization and optimization of image sensing toward continuous mobile vision. Mobisys '13, pages 69–82. ACM, 2013.

[21] M. MacKay. Camera lens and filter adapter assembly, May 4 1993. US Patent 5,208,624.

[22] N. Rajagopal, P. Lazik, and A. Rowe. Visual light landmarks for mobile devices. In *Proceedings of the 13th international symposium on Information processing in sensor networks*, pages 249–260. IEEE Press, 2014.

[23] S. Schmid, G. Corbellini, S. Mangold, and T. R. Gross. Led-to-led visible light communication networks. MobiHoc '13, pages 1–10. ACM, 2013.

[24] E. Star. *ENERGY STAR Program Requirements for Residential Light Fixtures*. 2010.

[25] X. Tang, Z. Ghassemlooy, S. Rajbhandari, W. Popoola, C. Lee, E. Leitgeb, and V. Ahmadi. Free-space optical communication employing polarization shift keying coherent modulation in atmospheric turbulence channel. In *Communication Systems Networks and Digital Signal Processing (CSNDSP), 2010 7th International Symposium on*, pages 615–620, July 2010.

[26] H.-M. Tsai, H.-M. Lin, and H.-Y. Lee. Demo: Rollinglight - universal camera communications for single led. MobiCom '14, pages 317–320, New York, NY, USA, 2014. ACM.

[27] A. Wang, S. Ma, C. Hu, J. Huai, C. Peng, and G. Shen. Enhancing reliability to boost the throughput over screen-camera links. MobiCom'14, pages 41–52. ACM, 2014.

[28] Z. Yang, X. Feng, and Q. Zhang. Adometer: Push the limit of pedestrian indoor localization through cooperation. *Mobile Computing, IEEE Transactions on*, 13(11):2473–2483, Nov 2014.

[29] Z. Yang, Z. Wang, J. Zhang, C. Huang, and Q. Zhang. Wearables can afford: Light-weight indoor positioning with visible light. In *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, pages 317–330. ACM, 2015.

[30] P. Yeh and C. Gu. *Optics of Liquid Crystal Displays*. Wiley Publishing, 2nd edition, 2009.

[31] C. Zhang and X. Zhang. Litell: robust indoor localization using unmodified light fixtures. pages 230–242, 2016.

[32] C. Zhang and X. Zhang. Pulsar: Towards ubiquitous visible light localization. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking*, pages 208–221. ACM, 2017.

[33] S. Zhu and X. Zhang. Enabling high-precision visible light localization in today's buildings. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 96–108. ACM, 2017.