

HoloCube: 3D Optical IoT Connections via Software Defined Pepper's Ghost

Xiao Zhang¹, Li Xiao², Matt W. Mutka²

University of Michigan-Dearborn¹, Michigan State University²
zhanxiao@umich.edu¹, lxiao@cse.msu.edu², mutka@msu.edu²

Abstract—Optical wireless communication (OWC) has inherent location aware and spatial reuse advantages over RF-based technologies due to the Line-of-Sight (LoS) propagation of optical signals. Hence, OWC presents a fresh opportunity for effective and secure IoT connectivity and data transmission. However, most OWC systems design transmitter as a point source without considering its spatial diversity in data delivery in 3D space, which is not suitable for real-world mobile IoT connections among devices and users. In this paper, we design and implement HoloCube, which provides 3D optical IoT connections via software defined optical camera communication and Pepper's ghost effect. At the heart of the HoloCube design is multiple virtual 3D hollowed-out cubes with adaptive Spatial-Color Shift Keying (SCSK) modulation. Specifically, the virtual cube seen from various directions has constant structure but embeds different data over time. The cube's positioning elements provide double reference for both 3D reconstruction (spatial) and robust color decoding (spectral). Our comprehensive experiments demonstrate that HoloCube achieves practical 3D omnidirectional IoT connections with 70 Kbps goodput at 4m in real-world indoor setting.

I. INTRODUCTION

With the rapid advancement of wireless technology, integrated communication and sensing (ICS) technologies have attracted significant interest from industry and academia. ICS is a promising development for next-generation wireless networks and IoT applications. These ubiquitous wireless connections aim to provide high-speed, ultra-low latency communication with enhanced security and precise sensing and localization capabilities. Consequently, ICS services are expected to boost broad future applications, including AR/VR/MR, and networks for vehicles and drones [1]–[9].

Among various ICS wireless methods, optical wireless technology which uses energy-efficient transceivers such as LED bulbs, screens, displays, and cameras, shows great potentials [1]. Optical signals possess a spectrum bandwidth 10,000 times broader than RF (radio frequency) wireless technologies like LTE, 5G, WiFi, and even high-frequency spectrums such as mmWave and THz [1], [10], [11]. Hence, optical wireless communication (OWC) faces no bandwidth congestion issues. Unlike RF signals, which propagate in Non-Line-of-Sight (NLoS) mode, optical signals travel directly in Line-of-Sight (LoS) manner. The LoS signal propagation in OWC offers three main advantages: (1) it establishes secure wireless connections within physical spaces, making eavesdropping difficult [3]; (2) it allows for numerous optical links to coexist, enabling practical IoT connections with spatial multiplexing

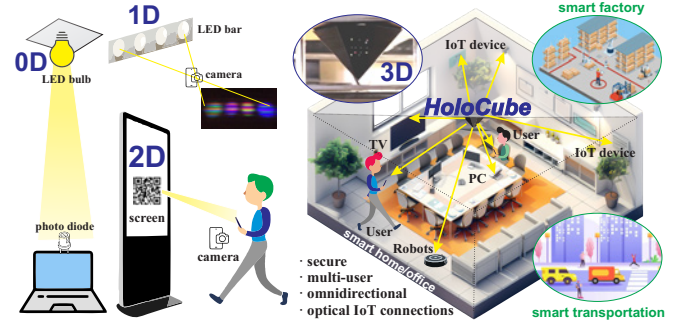


Fig. 1: Traditional optical wireless systems lack full 3D mobility, unlike HoloCube for 3D Optical IoT connections, which provides software defined integrated sensing and communication for indoor personal area network.

without significant interference [12]; and (3) it provides 3D location awareness in a visible and straightforward manner without disrupting communication services [13]–[15].

However, optical signals' LoS paths are prone to be blocked and require precise pointing between the transmitter and the receiver. For example, LiFi systems need the receiver to point (a single-pixel photo diode) to the transmitter (a single LED bulb) for successful communication [16], [17]. To extend users' mobility, we can adopt cameras as OWC receivers that consist of millions of pixels for light perception. For example, we can hold a camera to sense whether a light bulb is On/Off or scan a bar/QR code on a display within 180° instead of precise pointing needed for LiFi usage [18]. Even so, the IoT devices' connections are still restricted without full 3D directions and space, unlike RF wireless technologies. We found this limitation comes from non-3D optical transmitters, especially directional designs like LED bulbs (0D), linear LED bars (1D), or screens (2D), as shown in Figure 1.

To address the limitation above, we can design and use 3D optical transmitter instead of a point light source for integrated localization and omnidirectional communication for multiple IoT devices and users. As demonstrated in U-Star [13], the authors design a physically 3D hollowed-out cube-shaped passive optical tag and then achieve omnidirectional scanning via a camera in an underwater 3D space. Given that U-Star tags are passive optical tags with fixed data embedding, they cannot be used for active optical wireless transmission where the data varies over time. To enable *omnidirectional optical wireless links* for multiple IoT devices and users, we design a *3D-shaped* and *active* optical transmitter while using

commercial cameras as receivers.

This paper presents **HoloCube**, the virtual but visual 3D tags for omnidirectional optical wireless to break the bottleneck that exists for OWC approaches. HoloCube is a software-defined optical transmitter comprising an optical front-end (8-directional reflector and smartphone screen), data plane (frame generator on a smartphone), and control plane (via an inaudible channel). Users and IoT devices at different directions and locations in the same space can receive specific data simultaneously via handheld smartphones, even while moving. To reliably achieve 3D optical IoT connections, HoloCube has to tackle several critical technical challenges proposed by the omnidirectional service design goals:

- *3D Directed Optical Transmission.* Instead of using multiple LED bulbs oriented in different directions, we create software-defined optical virtual 3D tags containing directional data for IoT devices. Generating synchronized yet direction-specific optical signals in one transmitter for multiple users at different locations poses a challenge.
- *Robust Optical Decoding.* To achieve reliable optical wireless connections with adaptive data rate, the HoloCube exploits spectrum (colors) and spatial (spatial dots) diversities for data embedding that varies with image frames. It is difficult to differentiate these colorful dots at variable distances and ambient light.
- *Location-aware User Access.* The HoloCube is designed for omnidirectional optical delivery to cater to multiple users positioned in various directions and distances from the HoloCube. Identifying and serving multiple users while accounting for their specific locations and resource requirements presents a significant challenge.

Importantly, our proposed HoloCube enabled 3D optical IoT connections are generally applicable to various IoT applications and scenarios, which is essential for our motivation for next-gen IoT era. We demonstrate the applicability with the low-cost prototype and real-world experiments while they can be upgraded and extended for more applications.

In summary, our intellectual contributions are as follows:

- HoloCube is the *first* work to exploit Pepper's Ghost effect to generate virtual but real 3D tags for omnidirectional optical wireless connections.
- Our design extends optical camera communication distance from $<1\text{m}$ to 4m and angles from $<180^\circ$ to 360° with robust optical decoding via double references.
- We implement an inaudible control channel with a location-aware MAC design for adaptive resource allocation, multi user access and services.
- The comprehensive experiments show that our prototype achieves 3D connections with 70Kbps up to 4m in office setting and 100% accuracy of user identification.

The proposed HoloCube is poised to drive real-world applications utilizing 3D optical IoT connections, particularly in scenarios where radio frequency (RF) signals are limited. For instance, in medical settings, HoloCube can transmit data without interfering with sensitive equipment such as pacemakers,

implantable defibrillators, ventilators, and MRI machines. In smart homes, HoloCube can manage devices, including sensors in densely populated areas. In industrial environment, it can ensure reliable data transmission unaffected by electromagnetic noise from machinery and provides accurate location tracking (e.g., indoor traffic light system for robots and ground vehicles in warehouse or plants). Additionally, in high-security environments (e.g., government offices), it offers secure communication with minimal risk of interception. These benefits give HoloCube practical implications, making it especially effective in scenarios where RF technology may fall short.

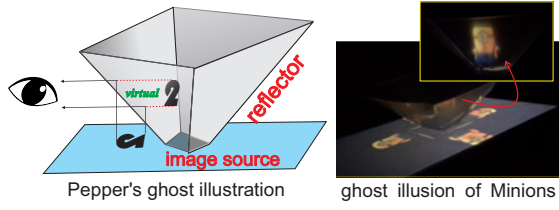
II. MOTIVATION

Drawbacks of Current OCC: Optical Wireless Communication (OWC) includes techniques like Li-Fi, Visible Light Communication, Optical Camera Communication (OCC), and Free Space Optical Communication (FSOC) [1]. Thanks to the popularity of smartphones equipped with cameras, OCC obtains great attentions for IoT connections. Despite advancements in camera imaging, existing OCC systems treat the LED bulb and camera as a point-to-point optical imaging system without spatial diversity. While researchers have increased the camera's light sensing frequency using the rolling shutter effect, it still lags behind the fast switching ability of LED bulbs, limiting OCC data rates to about 8 Kbps [3], [11], [19]–[23]. To enhance OCC, researchers use 2D spatial diversity in camera imaging by employing a commercial screen or display as an optical modulator [12], [24]. Unlike traditional point-to-point OCC, this approach operates at the frame rate level rather than the faster shutter rate. By utilizing millions of pixels in both the display and the camera's image sensor, data rates of about 1 Mbps can be achieved, as demonstrated in Aircode [12]. However, these screen-camera communications have limited working range (less than 1m) and scanning angles (less than 180°), making them impractical for omnidirectional optical IoT connections among multiple devices. Additionally, compared to 3D optical tags [13], 2D data embedding suffers from limited symbol distance for robust decoding.

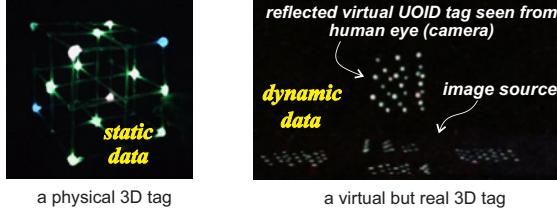
Software Defined Pepper's Ghost: Pepper's Ghost is a classic technique for creating transparent ghostly images [25]. Popularized by John Pepper in the 1800s, this technique reflects an image off plexiglass. It is commonly used in theaters, where viewers see a virtual image with depth, seemingly emerging from nowhere. Achieved by reflecting an image onto a transparent screen at a 45° angle, it creates a tangible object, as shown in Figure 2 (a) with example of ghost Minions.

There are three possible approaches to design a 3D omnidirectional optical transmitters: (1) design a 3D hollowed-out light source that consists of LED nodes controlled by a microcontroller, (2) physically holographic: generating the real 3D object in air, (3) visually holographic: displaying the virtual-3D object in 2D planes via the Pepper's ghost effect [26]. The first solution requires complex hardware implementation and control and cannot provide direction-specific data embedding. The techniques for (2) are being explored and the devices are very expensive [27]. HoloCube explored the third solution,

which is cost-effective with software defined optical potential and configuration flexibility.



(a) Pepper's ghost with Smartphone



a physical 3D tag
(b) HoloCube delivers dynamic data via Pepper's ghost

Fig. 2: (a) A smartphone with a Pepper's ghost reflector can show virtual 3D object. (b) HoloCube is software-defined and offers 3D optical connections for IoT devices.

III. HOLOCUBE IN A NUTSHELL

Overview. HoloCube facilitates 3D omnidirectional optical IoT connections among HoloCube tag and stationary or mobile IoT devices with camera as reader. It provides software-defined resource allocation, multiple user access with location awareness, and optical/audio dual-channel for data and control planes. For example, if an IoT device, such as a ground vehicle with a camera automatically oriented towards HoloCube (similar to a sunflower facing the sun), moves within an office, it can either receive data from HoloCube tag via optical IoT connection or report sensing data to HoloCube tag through an inaudible audio channel. Subsequently, HoloCube can deliver the requested data or control commands to specific IoT devices with adaptive data rate and brightness energy.

Applicable Scenarios. HoloCube is applicable in a wide range of scenarios since modules such as screens, cameras, speaker and microphone have been ubiquitously deployed as default on many edge devices [28]–[30]. HoloCube is designed to cater to various applications such as smart home, smart factory, and indoor navigation, offering secure IoT connections within spaces, free from eavesdropping risks. With upgrades, HoloCube can handle outdoor challenges, extending its use to smart transportation and vehicular networks.

System Architecture. HoloCube comprises two components, as illustrated in Figure 3: (1) 3D active virtual optical tags, referred to as HoloCube tag, and (2) AI-based mobile tag readers, which are massive IoT devices.

- **HoloCube Tags (Sec. IV):** Mounted on ceilings with fixed orientations, these tags include an optical front-end (a smartphone screen and a Pepper Ghost Reflector) and connect to the Internet, functioning as optical routers. They display eight virtual 3D cubes with consistent vertex colors (red, green, blue, and white) and unique data

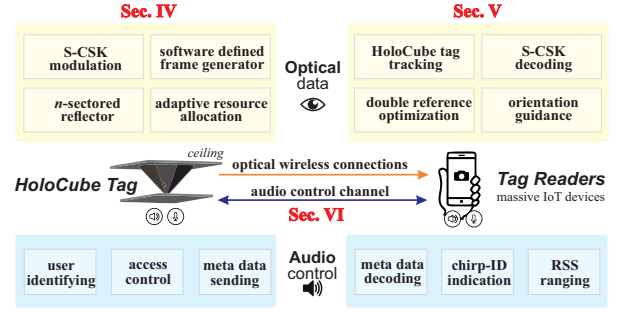


Fig. 3: HoloCube System Architecture.

elements. Proper spacing ensures clear data presentation by preventing Line-of-Sight (LoS) blockage.

- **Tag Readers (Sec. V):** Smart devices with cameras capture HoloCube images. They track tags with YOLO, classify visual orientation, estimate distance, and parse data using spatial and spectrum references to counter ambient light. *Spatial reference* restores 3D structure and locates data elements. *Spectrum reference* retrieves data by comparing colors to positioning elements.
- **Optical and Inaudible Dual Channels (Sec. VI):** We use spatial and spectrum diversity in optical imaging for rapid omnidirectional delivery in the optical data channel, employing techniques like S-CSK modulation and adaptive resource allocation. To ensure smooth user access and services, we also establish an inaudible control channel in the near-ultrasound spectrum, handling user identification, access control, and metadata transmission.

IV. TAG DESIGN

Three components make up the HoloCube tag we created: (1) a **video frame generator**, (2) an **omnidirectional reflector**, and (3) the **virtual yet real 3D cubes** that are the reflected images from the screen through a certain sector of the reflector. In the design of the HoloCube, there are some key parameters illustrated in Table I. S_i denotes the sector index with range from 1 to 8. O denotes the modulation order with the range of [3, 6]. θ denotes the angle between the sector plane and the screen plane in the range of [30°, 45°, 60°]. We set 45° in the final design with analysis. C denotes the divide color scales in the modulation. D denotes the transmission distance in the range of 1m to 5m. The 3D virtual optical cube is the de facto transmission front-end and we illustrate it first.

TABLE I: 5 key parameters in HoloCube tag design.

symbol	parameter	range	definition
S_i	sector index	[1,2,...,7,8]	the index number of sectors
O	modulation order	[3,4,5,6]	modulation order of 3D virtual cube
θ	reflection angle	[30°,45°,60°]	angle between sector and screen
N	color scales	[4, 8, 16]	the divided color scales in modulation
D	distance	[1m,...,5m]	the distance from user to HoloCube

A. 3D Virtual Optical Cube

3D Embedding Design. The presented 2D image of the 3D virtual cube seen from specific directions are expected to be different based on the perspective principle but they represent the exact one spatially-united (i.e., the same size and

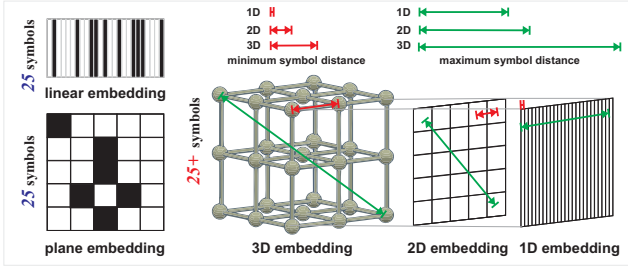


Fig. 4: 3D data embedding and symbol distance comparison with 1D and 2D optical codes.

orientation) 3D cube, as illustrated in Figure 2 (b). Given an example, a 3-order 3D cube has $3^3 = 27$ elements, which includes 8 vertices as positioning elements and the inside remaining 19 elements as data elements. These elements are spatially distributed in a 3D space instead of an existing bar/QR code where elements are distributed in a linear or planar manner with limited symbol distance and the scanning angle limitation. For instance, a data element (i.e., dot) in a cube has a broader normalized symbol distance distribution than a dot in a square, as illustrated in Figure 4. Besides, there is no LoS blockage issue in U-Star [13] thanks to our software defined video frame generation with preset proper viewing angles of the camera.

S-CSK modulation. In contrast to the static data in an U-Star tag [13], the color of each element in our virtual 3D cube can be set synchronously and separately to the expected color to represent the time-varied data. To boost the data rate, we adopt the Color Shift Keying (CSK) modulation in numerous spatial data elements, which is named as S-(N)-CSK modulation. N is the color scales listed in Tab I. Each data element can be the combination of different amounts of R (red), G (green), and B (blue) colors. Each pure color (R_p , G_p , B_p) is divided into N scales of $[0, 255]$. For example, if the N is 16. Then a data element's color mixed with Red of 240 (14×16), Green of 128 (07×16), and Blue of 48 (02×16), as shown in Figure 5. We define the $14 \times$, $07 \times$, and $02 \times$ as the α , β , and γ , formulated as below.

$$\mathcal{C} = \alpha \times R_p/16 + \beta \times G_p/16 + \gamma \times B_p/16 \quad (1)$$

Thus, each data element can denote $\log_2(16 \times 16 \times 16) = 12$ bits. A 3-order virtual 3D tag can embed $12 \times 19 = 228$ bits in one frame duration, and $228 \text{ bits} \times 60\text{Hz} \times 8 \text{ directions}$

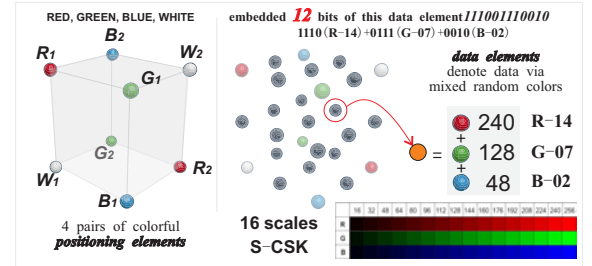


Fig. 5: 4 positioning element pairs based S-16-CSK (Spatial-Color Shift Keying) modulation.

$= 109.4 \text{ Kbps}$ data rate with 60 Hz display refresh rate. The benefit of S-CSK compared with traditional CSK is its spatial multiplexing via these numerous data elements instead of solely one light source. Furthermore, these spatially located data elements have no mutual signal interference thanks to the millions of pixels of both screen and the camera to separate them via the pinhole imaging principle.

B. Video Frame Generator

Center-Bias Elements Generation. An object or a cube seen from different view angles are different as caused by the perspective principle. Our goal is to generate these different viewed images of the same cube at the same time on the screen. The challenge is to generate each element of these 8 cube images (i.e., eight 3-order cubes have $27 \times 8 = 216$ elements with different sizes and specified setting colors) considering different screen sizes. We design the center-bias elements generation algorithms, as illustrated in Figure 6. There are 6 steps: (1) use Unity 3D platform to create 3D cube with 8 vertices, (2) capture the 2D image of the tag seen from 8 directions, (3) generate a frame consists of these 8 captured 2D images at their specific sectors, (4) find the center point of the frame and measure the normalized bias from each vertex of a cube image to the center point in both x and y directions, (5) generate all the positioning elements of cubes from 8 directions based on the center-bias values whatever the screen size is, (6) based on these positioning elements, it is easy to find other elements from a cube according to the space geometry (e.g., middle point, three equal points, etc.).

Adaptive Cube Order. The data rate of HoloCube increases with the cube order O and we can design the HoloCube with an adaptive cube order from the basic 3-order to the 6-order for an adaptive data rate (i.e., **13.68** to **149.76 Kbps** for a

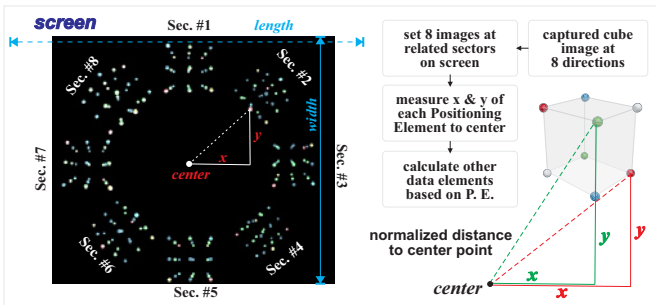


Fig. 6: 1 video frame with 8 3D cubes on sectors.

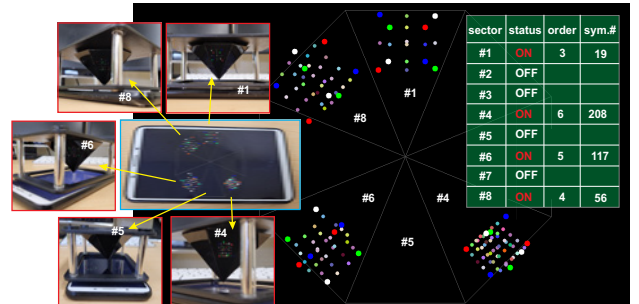


Fig. 7: One examples of 3D tags with adaptive orders.

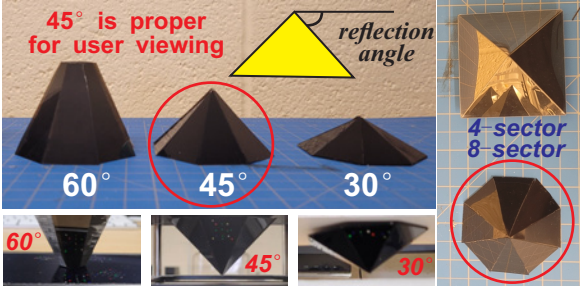


Fig. 8: 3D printed 45° reflector with 8 sectors.

cube with 60 Hz refresh rate and N is 16). If 8 directional cubes are with different data, the achieved data rate can vary from **109 Kbps** to **1.2 Mbps**. One example of cubes with adaptive orders is shown in Figure 7. To achieve these adaptive cube order, we can follow the above center-bias elements generation algorithms to first find positioning elements, then calculate the locations of other elements one by one. Although each element is calculated, the entire computation latency is less than the refresh period and therefore keep these elements shown and updated synchronously. The frame generating time grows exponentially as the S-CSK order increases. However, this latency can be reduced by searching for and displaying pre-generated frames from a database on a smartphone (i.e., using additional storage space to reduce time costs).

C. Omnidirectional Reflector

Spatial-United Cubes. The generated image frame on the screen is reflected by the omnidirectional reflector shaped with an 8-sector pyramid based on Pepper's Ghost illusion. The reflector is 3D printed with 8 sectors and the reflection angle is set as 45°. Each sector of the reflector is coated with black plastic thin film with a smooth plane. Thus the 8 virtual cubes on the screen can be seen by user via related sector of the reflector. Although these virtual cubes are a 2D image, they simulate a physical 3D cube in space, as illustrated in Figure 8 and Figure 9. If the inside data elements from these 8 cubes are the same (both spatial-united and data-united), it indeed is the same scene for the user's vision as a physical cube.

Direction-Specified Data. As shown in Figure 8, we print reflectors with 30°, 45°, and 60° reflection angles. The virtual

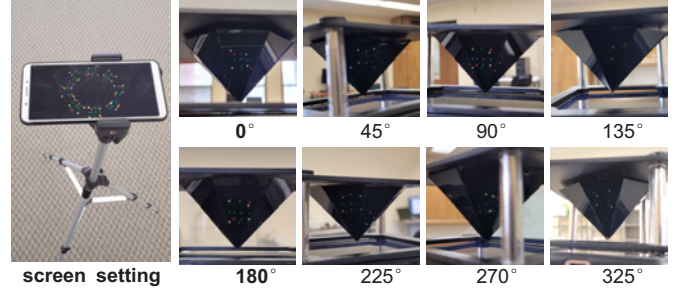


Fig. 9: Spatial-united cubes with direction-specific data.

cube generated by 60° reflector is too low to view while 30° reflector is too high to view. Thus, we choose the reflector with 45°. Besides the serving for single user, we should also consider serving multiple users in different directions to the HoloCube. Therefore, we design the reflector with 8 sectors instead of 4 is to fully utilize the finer-grained spatial diversity for multi-user services.

Thanks to the software defined optical feature of our HoloCube via the video frame generation, we can easily achieve a spatial-united cube but direction-specified data embedding and resource allocation for which a physical cube cannot achieve easily with a low cost. As shown in Figure 7 and Figure 9, we can set the 8 cubes with different embedded data for users at different directions. Besides the 8 sectors with 8 cubes for 8 directions, we can also conduct 4, 6, or other numbers of directions with simply an adaptation of the video frame generator and the reflector design.

V. READER DESIGN

In this section, we will focus on tag reader design. To decode the embedded data in a virtual but visual 3D cube inside of the HoloCube, the first task for the reader is to determine the location of the HoloCube from the surroundings. Then, the reader can further crop the 2D image of the inside virtual but visual 3D cube from the tracked HoloCube for next step: decoding. Convolutional Neural Networks (CNN) are widely applied in many applications especially in computer vision areas including image recognition, objectives tracking, image segmentation, etc. YOLO (You Only Look Once) models are widely used for objective detection due to their fast

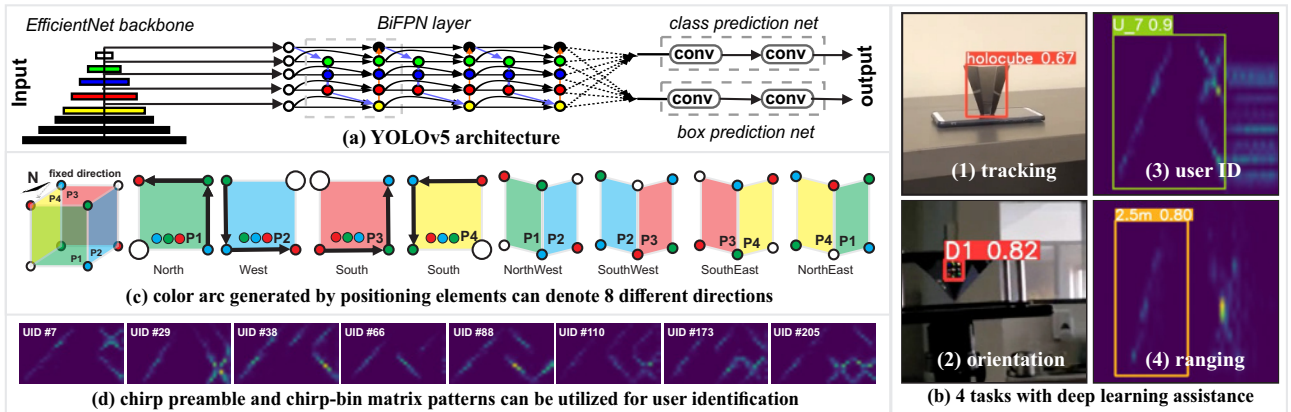


Fig. 10: YOLO assisted tag tracking, orientation guidance, and audio based user identification and ranging.

inference and robust performance. Thus, we adopt YOLOv5 model for 4 goals, as shown in Figure 10 (b): (1) real-time tag tracking, (2) user orientation determination, (3) user identification from spectrogram of chirp-bin matrix and (4) received signal strength based ranging.

A. Tag Tracking & Orientation Awareness

Tag Tracking via YOLO. As shown in Figure 10 (a), the adopted open sourced YOLOv5 CNN network consists of EfficientNet backbone, BiFPN layers for efficiently extract objectives' features. Then, these extracted features are fed into the prediction nets for HoloCube recognition with the bonding box (location determination). We captured the image of a HoloCube tag with random angles and varied distances within 5m at different altitudes. The total number of images for training is 100. We manually label the HoloCube tag with bonding box in images via Roboflow platform with data augmentation to increase the dataset size to 145 images.

User Orientation Determination. In our HoloCube design illustrated in Section IV-C, the data are directional for users at different directions, thus the reader should know the user's orientation based on the positioning elements (P.E.) seen from the virtual but visual 3D cube. The positioning elements are 4 pairs of colorful diagonal vertices. Thus, they can generate specific color arcs in each specific plane of the virtual 3D cube and thus to reflect user's orientation, as illustrated in Figure 10 (c). For example, the HoloCube is mounted in the ceiling with fixed orientation and the plane with anticlockwise colored-arc of white, blue, green and red is oriented to the South. If the reader captured this plane, then it reflects that the user is facing to the North. We also adopt the YOLO model for the eight orientation determination. We collect 20 images of HoloCube at each direction at 2 m and label them manually.

B. Data Parsing via Double Reference

Each data element in the virtual 3D cube contains the information of (1) where its **spatial** location is, and (2) what color **spectrum** (or light wavelength) it is. (1) is essential to recover the bit sequences in the order. (2) is essential to correctly parse the embedded bits in that data element. In our approach, we design **Double Reference** in spatial and spectrum aspects for robust data parsing at reader side.

3D Restoring via Positioning Elements. As shown in Figure 11 (a), the captured virtual 3D cube has 4 diagonal vertices pairs with pure red, green, blue and white colors. We

can filter out these vertices and then restore the 3D structure of the virtual cube from its 2D image based on the space geometry principle. The detailed steps of 3D restoring via positioning elements are: **S1**: Filter out 8 positioning elements from the captured 2D image, **S2**: Locate each data element in each layer, **S3**: Know the data elements sequences for potential bits decoding and re-connection. Given an example in Figure 11 (a), we can first reconstruct 3D structure and the 3D location of each data element based on positioning elements. Then we can detect correctly via the color reference and further parse the embedded 12 bits in each data element based on S-CSK demodulation. Finally, the embedded bit stream can be parsed and re-connected according to their embedding sequence of 3D locations.

Color Reference via Positioning Elements. Different from the static and sole color for data elements in U-Star tags (i.e., fixed dark or green) [13], the color of each data element in the virtual 3D cube is the combination of R,G,B color with 16 scales separately. These combined colors face severe distortion and attenuation after the propagation with the impact of distance and ambient light. Specifically, it is modeled in theory that the variation trend of R, G, and B ideally are different with individual non-linear color-distance function [31], as shown in Figure 11 (b). For example, the detected color of a data element at 0m is RGB (255,255,198). Nevertheless, with the increased distance from the camera to the HoloCube tag, all R, G, B color component values are decreased, the detected RGB value at 1m is (203,236,189) at 1m and the detected RGB value at 2m is (128,133,126), as shown in Figure 11 (c).

However, the color distortion has no unified function with distance under the varied optical environment, which makes the recovering of original color impossible based on the known distance and color-distance function for further correct decoding. For example, the detected pure color of R,G,B at 1m is changed from (255, 255, 255) to (205, 235, 222), which is normalized as (0.8, 0.92, 0.87), as shown in Figure 11 (d). This received color is not matched with the distorted color in theory, as shown in Figure 11 (b).

To overcome the challenge above, we endow positioning elements (P.E.) with the second reference, color reference, for robust decoding. Although the color distortion under varied distance and ambient light are not united, the color attenuation and distortion for all data elements and positioning elements in the virtual 3D tag are consistent because they nearly have

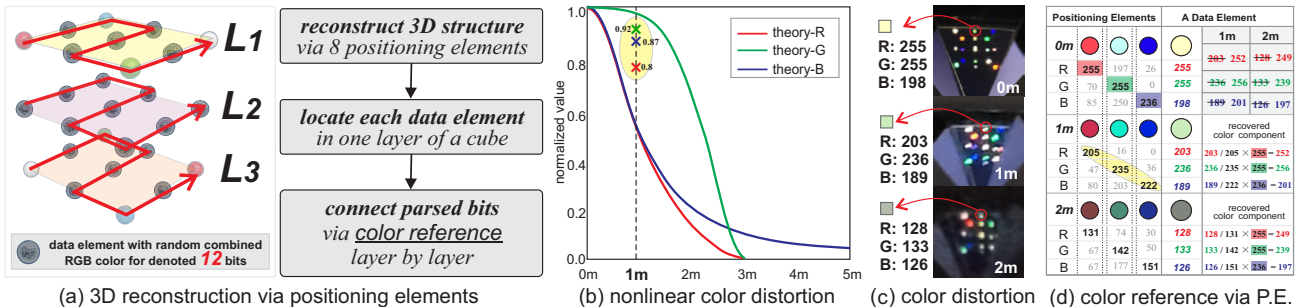


Fig. 11: P.E. based double references: (1) 3D reconstruction and (2) embedded data parsing via color reference.

the same propagation distance to the camera and ambient light impact. Thus, we can use positioning elements and their pure colors' variation to record the color distortion at that moment and location as the reference for other data elements. For example, as shown in Figure 11 (d), the received colors are successfully recovered to original transmitted colors via P.E. referred ratio (e.g., the B value 189 of data element to the B value 222 of blue positioning element at 1m).

VI. MAC DESIGN

To enable multi-user services, we've designed a location-aware MAC for HoloCube, operating inaudibly to avoid interference with the optical data plane and ambient sounds. The inaudible control channel, in the near ultrasound band, is enabled by our audio packet design, suitable for synchronous functions. Each audio packet includes chirp preambles for synchronization, distance ranging, and user indication, along with OFDM frames for metadata transmission. This design ensures short packet durations to periodically update user information to the HoloCube tag. In our design, we've set the audio packet duration to 200ms (5Hz) to accommodate potential long-distance ranging (e.g., $50\text{ms} \times 340\text{m/s} = 170\text{m}$). However, achieving all expected functions within this short inaudible packet is challenging.

A. Chirp based Preamble

Our designed chirp preamble consists of two parts: (1) an FMCW (Frequency Modulated Continuous Wave) up-chirp for synchronization and distance estimation, and (2) chirp-bin matrix to indicate and differentiate a large number of users.

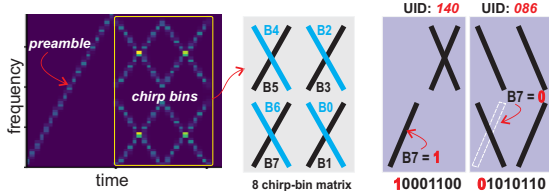


Fig. 12: Chirp based preamble consists of two segments: (1) FMCW up-chirp, and (2) chirp-bin matrix.

FMCW for Synchronization. In the control channel, the transmitter is the speaker in user's smartphone. The receiver is the microphone at the HoloCube tag. The FMCW can assist the receiver to know when the audio packet starts and where is the data frame for further meta data decoding. The f_{min} and the f_{max} of the FMCW up-chirp is set to 17 KHz and 20 KHz separately to guarantee it is imperceptible by human. This broad Δf make it easier for robust identification by reader (i.e., the HoloCube's microphone) at long distance [32].

Chirp-bin Matrix for User Indication. In the design of chirp-bin matrix, we divide the Δf into 2 sub-band and the Δt into 2 sub-duration, as shown in Figure 12. The chirp can be up-chirp as well as down-chirp. Thus, there are 4 possible chirp-bin shapes (i.e., with/without up-chirp and with/without down-chirp) in each spectrogram (i.e., voicegram) bin and totally $4^4 = 256$ status to indicate up to 256 users at the same time. Then the receiver (i.e., microphone at HoloCube) can utilize its onboard computation resource to run YOLO model

assisted user identification based on the generated spectrogram chirp-bin matrix patterns.

ToF Ranging vs. Signal Strength based Ranging. As demonstrated in [33] that we can utilize ultrasound FMCW chirp signals for ToF (Time-of-Flight) fine-grained and continuous ranging. As shown in the left of Figure 13, the sent-out audio signals can be reflected by the sectors of the HoloCube tag. We are supposed to obtain the round-trip time of audio signals via FMCW chirp's shift on the spectrogram to calculate the distance. However, the smartphone's mic captured audio signals are not fine-grained to show the significant gap between sent and echo signals.

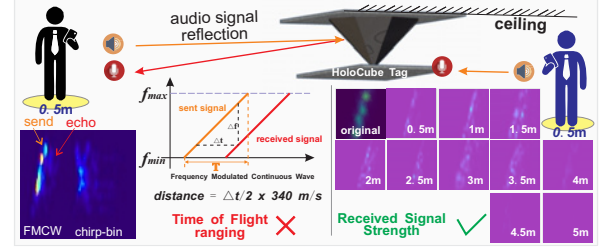


Fig. 13: ToF ranging vs. RSS ranging.

Therefore, we adopt the YOLO assisted received signal based distance estimation to provide user's localization for adaptive resource allocation, which requires the grain of the floor area for a human standing on (i.e., about 0.5m, a step length). As shown in the right of Figure 13, the user's speaker send out the audio packet at different distance to the HoloCube tag, the tag received audio signal strength is decreased by the increased distance. With YOLO deep learning assistance, we can classify the distances with 0.5m precision accordingly.

B. OFDM Meta Data

Following the chirp preamble, we design the OFDM (Orthogonal Frequency Division Multiplexing) symbol based segments for meta data transmission. The challenge here is to embed as much as data via the limited narrow bandwidth in near ultra sound band from 17KHz to 19.6KHz.

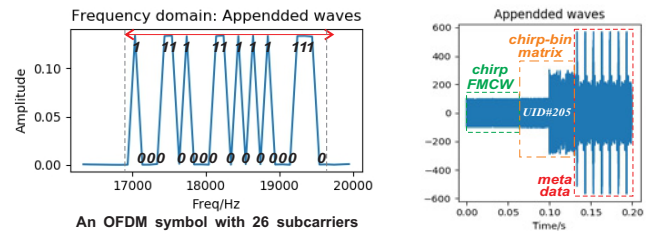
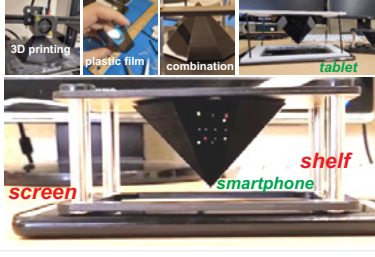


Fig. 14: An audio packet (200ms) with chirp preamble, chirp-bin matrix and OFDM data frame.

OFDM Symbol Design. We divide the near-ultra-sound audio band into 26 subcarriers with a gap of 100 Hz, as shown in Figure 14 (a). In order not to introduce additional decoding overhead, we adopt simple OOK (On-Off Keying) modulation. If the amplitude of the subcarrier is high, it denotes the bit 1 and otherwise the bit 0. Combined with a preamble, the entire audio packet has a duration of 200ms to update the calculated distance or meta data from the user side to the tag side 5 times per second periodically. **Meta Data Content.**



module	price	details
reflector	5	3D printed pyramid
film	1	attached plastic film
shelf	2	handcraft shelf structure
screen	≈ 50	reused/old smartphone
total	< 70	cost can be easily controlled

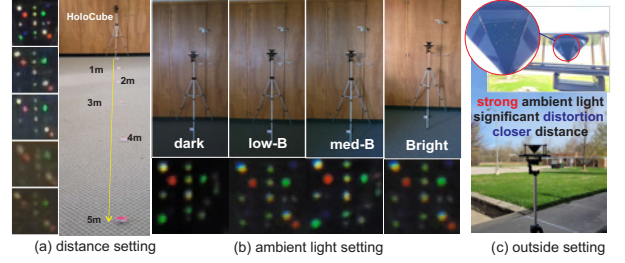


Fig. 15: HoloCube prototype. Fig. 16: Cost of HoloCube tag.

Fig. 17: Experiment setup and outside testing.

The inaudible channel can transmit a user's access request, location information, location switching, and other data to the HoloCube as an up-link, enabling adaptive multi-user services.

VII. EVALUATION

A. HoloCube Tags

We implement a proof-of-concept HoloCube tag based on commercial devices. As shown in Figure 15, the HoloCube prototype consists of a smartphone (i.e., VIVO-Y71) and a 3D-printed pyramid reflector with 8 sectors attached with black and smooth plastic paper. The screen size of smartphone is 720×1440 pixels of the 6.00-inch display. The reflector has the height of 7 cm and the width of 3.5 cm. The angle between a sector plane to the screen plane is 45° . The reflector is 50 g and attached on the screen with double-sides tap. The entire HoloCube can be mounted on the indoor ceiling or the top of other static/mobile objectives as omnidirectional optical transmitter. The entire cost of the HoloCube prototype is less than \$70 and the details are listed in Figure 16.

B. Tag Reader

The tag readers are the users' hand-held smartphones. These commercial smartphones have cameras, a microphone and a speaker. In our experiments, we use a commercial smartphone Samsung S20 for comprehensive experiments. Considering that the smartphones are already provided by users, the cost of the tag reader should not be counted into cost as the smartphone used in HoloCube tag. We use the smartphone to take images of HoloCube for further decoding and send out audio signals to the HoloCube. We present in this section our experimental setup for HoloCube evaluation and detailed performance of our proof-of-concept HoloCube prototype. We set different conditions for experiments including distances and ambient light settings. The experiment scenarios are shown in Figure 17. We evaluate the BER and related goodput based on the reflected image of HoloCube as color reference (0m, medium brightness). We also placed our current HoloCube prototype, which is designed for indoor use, in an outdoor environment to identify challenges and explore potential solutions.

C. Optical Link Performance

1) BER and throughput results.:

BER vs. different colors. The color components of R, G, and B have different non-linear distortion in propagation. Therefore, we evaluate our color reference performance for different color components with 8 color scales. As shown in Figure 18 (a), the BER of red and green increase significantly

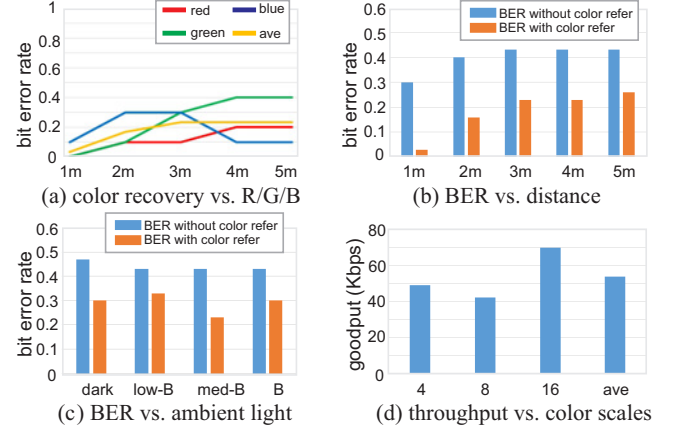


Fig. 18: BER results and related goodput performance.

after 3m while the BER of blue decreases after 3m, which is reasonable because the blue color has more recovery range compared with red and green, as shown in Figure 11 (b).

BER vs. different distances D . We set the distance between the reader (camera) and the HoloCube tag in range of [1m, 2m, 3m, 4m, 5m] and capture images of the 3-order virtual 3D cube with 8 color scales for each setting indoor. As shown in Figure 18 (b), the BER increases with the increased transmission distance from 0.03 at 1m to 0.26 at 5m. All the BER with color reference are lower significantly than without color reference in decoding, which keep high in all distances.

BER vs. different ambient light. We set the distance at 2m with 4 ambient light settings [dark, low-brightness, medium-brightness, bright, strong-brightness] to capture images of 3-order virtual 3D cube with 8 color scales for each setting. As shown in Figure 18 (c), BER with color reference drops compared to decoding without color reference, and there is no significant difference across ambient light settings thanks to the color reference from positioning elements. Outdoor testing (Fig. 17 (c)) revealed that very strong ambient light (i.e., sunlight) mostly destroys cube detection and further decoding, unless proper parameters are used (e.g., $D < 0.2m$, camera exposure settings).

Throughput vs. different color scales N . We set the HoloCube order as 3 (i.e., the $3 \times 3 \times 3$ cube) and test different color scale division in modulation of [4, 8, 16] (i.e., divide pure color into 4, 8, and 16 scales to denote bits, as shown in Figure 5) impact of obtained goodput at 4m under bright ambient light and 60FPS screen refresh rate setting. As shown in Figure 18 (d), 16 color scales achieves the highest data rate of 70 Kbps. Although the higher color scales cause the

decreased symbol distance and higher BER, the embedded bits in one data element is 12 ($3 \times \log_2 16$) which is more than 6 bits in 4-color-scale and 9 bits in 8-color-scale setting.



Fig. 21: Tracking evaluation setting and examples.

2) **Tag Tracking:** The trained model achieves precision of 0.947 and the recall of 0.95. As shown in Figure 19 (a), we evaluated 4 types of motion at 5m: (1) horizontal translation motion, (2) vertical translation motion, (3) anteroposterior translation motion, and (4) encircling motion, as shown in Figure 21. The tag recognition accuracy and the latency is important for user experience and system robustness. **Tracking Accuracy.** As shown in Figure 19 (b), the tracking of HoloCube tag is accurate at 5m whatever the motion is. All of the tracking accuracy is over 0.99. **Tracking Latency.** The tracked results should be updated in real time to guarantee the user experience. As shown in Figure 19 (b), there is not significant difference of tracking latency among different motions. The average of tracking latency is 21ms with min of 10ms. Thus, the screen refresh rate can set about 47FPS for further optical transmission.

3) **Orientation Guidance:** The 8 directions in orientation guidance are North, NorthEast, East, SouthEast, South, SouthWest, West, and NorthWest, labeled them as D1-D8. We first collect 6 images for each direction, totally 48 images and then perform dataset augmentation. There are 114 images in the orientation dataset after augmentation. We used the model after 300 epoches training for real-world evaluation. **Orientation Guidance Accuracy.** As shown in Figure 19 (c), the average of orientation of 8 directions achieves 0.91 and 6 of 8 directions achieves over 0.94. The accuracy drop of D5 (0.69) and D8 (0.8) is because of non-detections instead of wrong direction determination. **Orientation Guidance Latency.** As shown in Figure 19 (c), all 8 directions' orientation guidance are similar with the average latency of 8.4ms to 8.8ms.

4) **Comparison with 1D/2D Tags:** We also present 1D and 2D virtual codes with the same edge size of 1.5cm and the same amount of embedded data elements (i.e., 19) of a 3-order 3D virtual cube with the same manner of HoloCube for comparison, as shown in Figure 20. **Symbol Distance.** With the same edge size of 1.5cm, the average symbol distance from the first data element to other data elements in virtual 3D cube is 1.4cm, which is $2.45 \times$ and $1.82 \times$ to the 1D bar code (0.57cm) and 2D QR code (0.77cm) via its 3D spatial element

embedding instead of linear and plane manners, as illustrated and measured in Figure 20. **FOV.** Although bar/QR codes with our reflector can be scanned in all directions, they look like 8 codes in 8 planes (sectors) instead of HoloCube's 1 spatial-united virtual cube. **Data Rate/Throughput.** Compared to U-Star's static data (21 bits, 3rd order, On-Off keying) [13], HoloCube transmits temporal data at 70 Kbps with the same order and 16-S-CSK, achieving a $3333 \times$ higher data rate. These results demonstrate that HoloCube effectively extends symbol distance, enhances decoding robustness, and increases data capacity.

D. Control Channel Performance

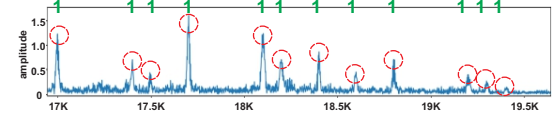


Fig. 22: One received OFDM symbol example at 5m.

1) Audio Link Performance:

We sent 5 audio packets (each packet has 26 embedded bits, illustrated in Figure 14) from the smartphone (user side) to the microphone (tag side) for each distance setting of [1m, 2m, 3m, 4m, 5m] with the same amplitude of inaudible signals. Then we process the received audio packets and decode the embedded bits based on the amplitude of OOK symbols carried by OFDM subcarriers. Figure 22 is a received OFDM packet example at 5m. As shown in Figure 23 (a), the BER keeps low among all distance settings. The average BER is 0.04 while the 1m achieves the lowest BER of 0.023 and 5m achieves the highest BER of 0.069.

2) **User Identification Performance: Setting.** Totally, we set 8 UID which are **7** (0b 00000111), **29** (0b 00011101), **38** (0b 00100110), **66** (0b 001000010), **88** (0b 01011000), **110** (0b 01101110), **173** (0b 10101101), and **205** (0b 11001101). We send out the audio chirp packets and receive them at random distance within 5m and convert it to spectrogram. Examples of spectrogram for each picked user are presented in Figure 10 (d). We captured 10 spectrogram for each user and totally 80 images. We also adopt augmentations with adjustment of grayscale, hue, saturation, and noise to increase the dataset size (192 images) and model robustness. **Results.** We use the trained YOLO model to parse user ID according to chirp-bin matrix pattern. Precision is the ratio of true positive samples to all samples predicted as positive by a classifier. Therefore, we use precision-confidence curve to evaluate the user identification performance. As shown in Figure 23 (b), the trained model successfully differentiates users with high

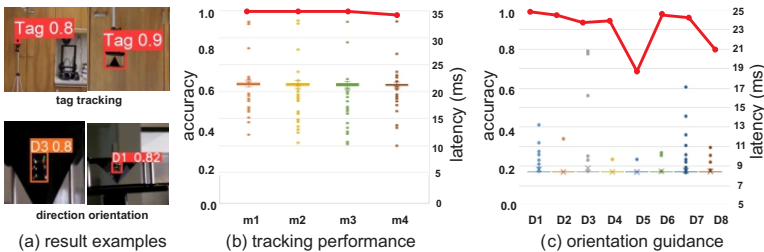


Fig. 19: Tag tracking and orientation guidance.

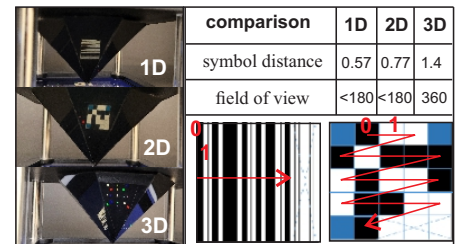


Fig. 20: Comparison with 1D and 2D codes.

confidence. All eight user classes achieve 100% prediction accuracy with confidence over 0.972.

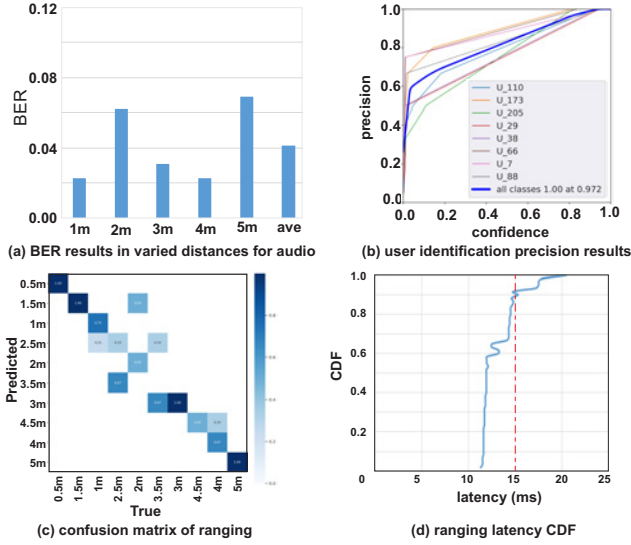


Fig. 23: Other evaluation results.

3) *Audio based Distance Estimation: Setting.* We capture 20 audio spectrogram images for 8 distance setting with augmentation as the dataset with 438 images, as shown in Figure 13. Then we use the trained model to parse distances from the user to the tag with 5 audio packets at each distance setting of 0.5m-5m with step of 0.5m. **Accuracy.** The confusion matrix of different ranging distances is shown in Figure 23 (c). Most of distance settings can be predicted correctly, especially 0.5m, 1.5m, 3m, and 5m. For other distance, the wrong predictions are still within the error of 0.5m-1m. These results demonstrate that our user distance ranging works well to provide the user's location with precision of 0.5m, a step size of a person. **Latency.** As shown in Figure 23 (d), more than 90% of distance ranging of a packet costs less than 15ms (i.e., 66.7Hz) while the maximum ranging latency is less than 22ms. These results show that our audio received signal strength based user distance ranging can work in real time.

VIII. RELATED WORK AND DISCUSSION

Related Work. Various communication technologies serve distinct purposes [34]–[39]. WiFi provides high-speed data transfer for HD video streaming and large file sharing, with rates ranging from hundreds of Mbps to several Gbps. Zigbee supports low-power IoT devices within shorter ranges, typically offering rates from tens to hundreds of kbps. Bluetooth facilitates short-range connections like wireless headphones and keyboards, with rates from tens of kbps to hundreds of Mbps. LoRa specializes in low-power, long-range IoT communications, with rates from hundreds of bps to tens of Kbps, suitable for smart cities and agricultural sensor networks.

In contrast, our HoloCube prototype with rates from tens of Kbps to several Mbps stands out for its ability for IoT connections. HoloCube provides moderate-capacity, low-latency, energy-efficient and highly secure communication, making it ideal for IoT applications requiring dependable, affordable

3D connectivity. Unlike RF technologies, which may face interference and eavesdropping risk, HoloCube offers a reliable and cost-effective solution for short-range 3D optical IoT connections in indoor environments like smart home/factory. Unlike Free Space Optical Communication (FSOC), which focuses on long-distance transmission of data, the Pepper's ghost effect primarily deals with short-range displays.

Discussion. Currently, our HoloCube system has limitations in handling information processing between adjacent sectors and seamless handover between sectors. Additionally, its performance is relatively low, with data rates around 70 Kbps (i.e., the similar rate level with LoRa) and a range of approximately 4 meters. Reliable demodulation is only achievable indoors [40], [41]. To enable HoloCube to operate in larger indoor and even outdoor environments, such as shopping malls or smart transportation facilities, the design and implementation of updated algorithm is required. **YOLO models:** We utilize YOLO models for complex localization tasks related to image patterns, rather than for communication tasks. The trained YOLO models (e.g., tag tracking, user orientation determination, user identification, and received signal strength-based ranging) can be deployed on mobile devices for real-world applications of the HoloCube system. These models can be integrated into commercial Android and iPhone devices using methods such as TorchScript, ONNX, and TensorFlow Lite [42]–[44]. **Vulnerabilities and solutions:** The Pepper's ghost effect is affected by ambient light, with high ambient light intensity can cause the image to become faint. This can be addressed by using a high-brightness projection source. Image quality can also be affected by interference from other light sources and the optical properties of the reflective surface. To mitigate these issues, signal strength can be increased by improving the optical system, reducing ambient light [45], and employing signal processing and machine learning-based denoising techniques such as GAN (Generative Adversarial Network) [46]. **Security and privacy concerns:** The MAC protocol ensures data access is restricted to authenticated users within the LoS space, providing dual-layer security protection.

IX. CONCLUSION

In this paper, we propose, design and implement the HoloCube system for omnidirectional software-defined optical wireless communication. We investigate software defined Pepper's ghost illusion using 3D spatial and spectrum diversities to enable omnidirectional IoT connections. We tackle challenges such as omnidirectional optical delivery, robust optical decoding, and location-aware MAC protocols. Lastly, we assess the performance of HoloCube in various scenarios. Our prototype, built with affordable commercial devices, achieves a goodput of up to 70 Kbps at a distance of 4m and provides real-time user distance parsing with a precision of 0.5m.

ACKNOWLEDGEMENT

We appreciate the constructive suggestions from the reviewers and Dr. Sangtae Ha for shepherding this paper. This work is supported by the National Science Foundation under Grant No. CNS-2226888 and Grant No. CCF-2007159.

REFERENCES

- [1] M. Z. Chowdhury, M. T. Hossan, A. Islam, and Y. M. Jang, "A comparative survey of optical wireless technologies: Architectures and applications," *IEEE Access*, vol. 6, pp. 9819–9840, 2018.
- [2] Y.-S. Kuo, P. Pannuto, K.-J. Hsiao, and P. Dutta, "Luxapose: Indoor positioning with mobile phones and visible light," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 447–458.
- [3] H. Pan, Y.-C. Chen, L. Yang, G. Xue, C.-W. You, and X. Ji, "mqrcode: Secure qr code using nonlinearity of spatial frequency in light," in *The 25th Annual International Conference on Mobile Computing and Networking*, 2019, pp. 1–18.
- [4] Z. Xiao and Y. Zeng, "An overview on integrated localization and communication towards 6g," *Science China Information Sciences*, vol. 65, pp. 1–46, 2022.
- [5] Y. Zhang, Y. Wang, L. Yang, M. Wang, Y.-C. Chen, L. Qiu, Y. Liu, G. Xue, and J. Yu, "Acoustic sensing and communication using meta-surface," in *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*, 2023, pp. 1359–1374.
- [6] E. Soltanaghaei, A. Prabhakara, A. Balanuta, M. Anderson, J. M. Rabaey, S. Kumar, and A. Rowe, "Proceedings of mobicom 2021," in *Proceedings of the annual International Conference on Mobile Computing and Networking*, 2022.
- [7] H. Rahmani, D. Shetty, M. Wagih, Y. Ghasempour, V. Palazzi, N. B. Carvalho, R. Correia, A. Costanzo, D. Vital, F. Alimenti *et al.*, "Next-generation iot devices: Sustainable eco-friendly manufacturing, energy harvesting, and wireless connectivity," *IEEE Journal of Microwaves*, vol. 3, no. 1, pp. 237–255, 2023.
- [8] X. Zhang, L. Xiao, and M. W. Mutka, "Poster: Holographic optical tag enabled omnidirectional iot connection," in *Proceedings of the ACM SIGCOMM 2024 Conference: Posters and Demos*, 2024, pp. 31–32.
- [9] X. Zhang, J. Mariani, L. Xiao, and M. W. Mutka, "Lifod: Lighting extra data via fine-grained owc dimming," in *2022 19th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON)*. IEEE, 2022, pp. 73–81.
- [10] P. Wang, L. Feng, G. Chen, C. Xu, Y. Wu, K. Xu, G. Shen, K. Du, G. Huang, and X. Liu, "Renovating road signs for infrastructure-to-vehicle networking: a visible light backscatter communication and networking approach," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking*, 2020, pp. 1–13.
- [11] Z. Tian, Y.-L. Wei, W.-N. Chang, X. Xiong, C. Zheng, H.-M. Tsai, K. C.-J. Lin, and X. Zhou, "Augmenting indoor inertial tracking with polarized light," in *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*, 2018, pp. 362–375.
- [12] K. Qian, Y. Lu, Z. Yang, K. Zhang, K. Huang, X. Cai, C. Wu, and Y. Liu, "Aircode: Hidden screen-camera communication on an invisible and inaudible dual channel," in *NSDI*, 2021, pp. 457–470.
- [13] X. Zhang, H. Guo, J. Mariani, and L. Xiao, "U-star: An underwater navigation system based on passive 3d optical identification tags," in *Proceedings of the 28th Annual International Conference on Mobile Computing And Networking*, 2022, pp. 648–660.
- [14] X. Zhang, G. Klevering, X. Lei, Y. Hu, L. Xiao, and G.-H. Tu, "The security in optical wireless communication: A survey," *ACM Computing Surveys*, vol. 55, no. 14s, pp. 1–36, 2023.
- [15] X. Zhang, G. Klevering, J. Wang, L. Xiao, and T. Li, "Rofin: 3d hand pose reconstructing via 2d rolling fingertips," in *Proceedings of the 21st Annual International Conference on Mobile Systems, Applications and Services*, 2023, pp. 330–342.
- [16] H. Wu, Q. Wang, J. Xiong, and M. Zuniga, "Smartvlc: When smart lighting meets vlc," in *Proceedings of the 13th International Conference on emerging Networking EXperiments and Technologies*, 2017, pp. 212–223.
- [17] X. Zhang and L. Xiao, "Lighting extra data via owc dimming," in *Proceedings of the Student Workshop*, 2020, pp. 29–30.
- [18] —, "Rainbowrow: Fast optical camera communication," in *2020 IEEE 28th International Conference on Network Protocols (ICNP)*. IEEE, 2020, pp. 1–6.
- [19] P. Hu, P. H. Pathak, X. Feng, H. Fu, and P. Mohapatra, "Colorbars: Increasing data rate of led-to-camera communication using color shift keying," in *proceedings of the 11th ACM conference on Emerging Networking experiments and technologies*, 2015, pp. 1–13.
- [20] X. Zhang, G. Klevering, and L. Xiao, "Posefly: On-site pose parsing of swarming drones via 4-in-1 optical camera communication," in *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2023, pp. 67–76.
- [21] X. Zhang, G. Klevering, K. Wijewardena, and L. Xiao, "Integrated on-site localization and optical camera communication for drones," in *2023 IEEE 24th International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE, 2023, pp. 334–336.
- [22] X. Zhang, G. Klevering, J. Mariani, L. Xiao, and M. W. Mutka, "Boosting optical camera communication via 2d rolling blocks," in *2023 IEEE/ACM 31st International Symposium on Quality of Service (IWQoS)*. IEEE, 2023, pp. 1–4.
- [23] X. Zhang, *Exploring Spatial-Temporal Multi-Dimensions in Optical Wireless Communication and Sensing*. Michigan State University, 2023.
- [24] T. Li, C. An, X. Xiao, A. T. Campbell, and X. Zhou, "Real-time screen-camera communication behind any scene," in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, 2015, pp. 197–211.
- [25] J. H. Pepper, *True History of the Ghost: And All about Metempsychosis*. Cambridge University Press, 2012.
- [26] T. Conner, "Pepper's ghost and the augmented reality of modernity," *Journal of Science & Popular Culture*, vol. 3, no. 1, pp. 57–79, 2020.
- [27] Y. V. Miklyaev, D. C. Meisel, A. Blanco, G. von Freymann, K. Busch, W. Koch, C. Enkrich, M. Deubel, and M. Wegener, "Three-dimensional face-centered-cubic photonic crystal templates by laser holography: fabrication, optical characterization, and band-structure calculations," *Applied Physics Letters*, vol. 82, no. 8, pp. 1284–1286, 2003.
- [28] N. Bazhenov and D. Korzun, "Use of everyday mobile video cameras in iot applications," in *Conference of Open Innovations Association, FRUCT*, no. 22. FRUCT Oy, 2018, pp. 305–308.
- [29] D. J. Dubois, R. Kolcun, A. M. Mandalari, M. T. Paracha, D. Choffnes, and H. Haddadi, "When speakers are all ears: Characterizing misactivations of iot smart speakers," *Proceedings on Privacy Enhancing Technologies*, 2020.
- [30] X. Zhang, G. Klevering, X. Lei, Y. Hu, L. Xiao, and G.-H. Tu, "The security in optical wireless communication: A survey," vol. 55, no. 14s, 2023. [Online]. Available: <https://doi.org/10.1145/3594718>
- [31] T. Akenine-Mo, E. Haines, N. Hoffman *et al.*, "Real-time rendering," 2018.
- [32] T. Wang, D. Zhang, Y. Zheng, T. Gu, X. Zhou, and B. Dorizzi, "C-fmcw based contactless respiration detection using acoustic signal," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 1, no. 4, pp. 1–20, 2018.
- [33] S. Laureti, M. Mercuri, D. A. Hutchins, F. Crupi, and M. Ricci, "Modified fmcw scheme for improved ultrasonic positioning and ranging of unmanned ground vehicles at distances; 50 mm," *Sensors*, vol. 22, no. 24, p. 9899, 2022.
- [34] L. Chettri and R. Bera, "A comprehensive survey on internet of things (iot) toward 5g wireless systems," *IEEE Internet of Things Journal*, vol. 7, no. 1, pp. 16–32, 2019.
- [35] M. Elkhodr, S. Shahrestani, and H. Cheung, "Emerging wireless technologies in the internet of things: a comparative study," *arXiv preprint arXiv:1611.00861*, 2016.
- [36] Y. Wan, K. Xu, F. Wang, and G. Xue, "Iotathena: Unveiling iot device activities from network traffic," *IEEE Transactions on Wireless Communications*, vol. 21, no. 1, pp. 651–664, 2021.
- [37] J. Jin, L. Feng, J. Wang, D. Chen, and H. Lu, "Signature codes in visible light positioning," *IEEE Wireless Communications*, vol. 28, no. 5, pp. 178–184, 2021.
- [38] K. Cui, Q. Yang, Y. Zheng, and J. Han, "mmripple: Communicating with mmwave radars through smartphone vibration," in *Proceedings of the 22nd International Conference on Information Processing in Sensor Networks*, 2023, pp. 149–162.
- [39] Y. Li, Q. Li, Z. Zhang, G. Baig, L. Qiu, and S. Lu, "Beyond 5g: Reliable extreme mobility management," in *Proceedings of the Annual conference of the ACM Special Interest Group on Data Communication on the applications, technologies, architectures, and protocols for computer communication*, 2020, pp. 344–358.
- [40] J. Wang, X. Zhang, L. Xiao, and T. Li, "Survey for soil sensing with iot and traditional systems," *Network*, vol. 3, no. 4, pp. 482–501, 2023.
- [41] C. Li and Z. Cao, "Lora networking techniques for large-scale and long-term iot: A down-to-top survey," *ACM Computing Surveys (CSUR)*, vol. 55, no. 3, pp. 1–36, 2022.

- [42] ultralytics. (2024) Ultralytics yolo docs model deployment options. Accessed: 2024-09-08. [Online]. Available: <https://docs.ultralytics.com/guides/model-deployment-options/>
- [43] A. Shridhar, P. Tomson, and M. Innes, “Interoperating deep learning models with onnx. jl,” in *Proceedings of the JuliaCon Conferences*, vol. 1, no. 1, 2020, p. 59.
- [44] A. Abdelhamed, S. Lin, and M. S. Brown, “A high-quality denoising dataset for smartphone cameras,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1692–1700.
- [45] Q. N. Pham, V. P. Rachim, J. An, and W.-Y. Chung, “Ambient light rejection using a novel average voltage tracking in visible light communication system,” *Applied Sciences*, vol. 7, no. 7, p. 670, 2017.
- [46] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.