EchoSight: Streamlining Bidirectional Virtual-physical Interaction with In-situ Optical Tethering

Jingyu Li
Computer Science
Peking University
Beijing, China
motren909@pku.edu.cn

Qingwen Yang
Computer Science
Peking University
Beijing, China
yangqingwen@pku.edu.cn

Kenuo Xu
Computer Science
Peking University
Beijing, China
kenuo.xu@pku.edu.cn

Yang Zhang
Electrical and Computer Engineering
University of California, Los Angeles
Los Angeles, California, USA
yangzhang@ucla.edu

Chenren Xu
Computer Science
Peking University
Beijing, China
chenren.xu@gmail.com

Figure 1: EchoSight leverages in-situ optical backscatter tethering to achieve a look-and-control bidirectional interaction on commercial AR glasses. This process mimics human's natural interaction with low mental burden. EchoSight requires no pre-registration or network connection, facilitating scenarios where interactions are opportunistic and impromptu.

Abstract

Emerging AR applications require seamless integration of the virtual and physical worlds, which calls for tools that support both passive perception and active manipulation of the environment, enabling bidirectional interaction. We introduce EchoSight, a system for AR glasses that enables efficient look-and-control bidirectional interaction. EchoSight exploits optical wireless communication to instantaneously connect virtual data with its physical counterpart. EchoSight's unique dual-element optical design leverages beam directionality to automatically align the user's focus with target objects, reducing the overhead in both target identification and subsequent communication. This approach streamlines user interaction, reducing cognitive load and enhancing engagement. Our evaluations demonstrate EchoSight's effectiveness for room-scale communication, achieving distances up to 5 m and viewing angles up to 120 degrees. A study with 12 participants confirms EchoSight's improved efficiency and user experience over traditional methods, such as QR Code scanning and voice control, in AR IoT applications.

CHI '25, Yokohama, Japan

ACM ISBN 979-8-4007-1394-1/25/04
https://doi.org/10.1145/3706598.3713925

CCS Concepts

• **Human-centered computing** → **Mixed / augmented reality**.

Keywords

Augmented Reality, Bidirectional Interaction, Optical Wireless Communication, Intuitive Interface, Backscatters

ACM Reference Format:
Jingyu Li, Qingwen Yang, Kenuo Xu, Yang Zhang, and Chenren Xu. 2025. EchoSight: Streamlining Bidirectional Virtual-physical Interaction with In-situ Optical Tethering. In *CHI Conference on Human Factors in Computing Systems (CHI '25), April 26–May 01, 2025, Yokohama, Japan.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3713925

1 Introduction

The rapid advancement in Augmented Reality (AR) technology has significantly enhanced our ability to explore and interact with the environment. Traditionally, AR has focused on directing data flow from the physical world to the virtual realm, aiming to create digital counterparts that enhance human passive perception. We refer to these AR systems as *passive AR*, which primarily processes sensory data from the physical world, such as a user's location, orientation, and the surrounding physical anchors. Consequently, AR has led to wide applications in education [5, 72], entertainment [32, 50], and the retail industry [3, 49], where augmented perception improves efficiency and experience.

Although passive AR has profoundly influenced how we perceive the world, there is a growing need for *active AR* that enables the users to actively alter real-world objects' states, such as changing appearance or making control actions. Nowadays, researchers have been dedicated to creating systems that emphasize both passive perception and active manipulation capabilities. The recent proposal of bidirectional interaction [12, 38, 75] aims to seamlessly bridge virtual and physical realms by combining passive AR and active AR. It supports passive augmented environment perception, while also providing an active way to influence physical objects in the environment. For instance, Sketched Reality [29] and MechARspace [76] emphasize the active data path that alters the physical states of toys or robots to reflect the changes of virtual counterparts. InfoLED [68] and LightAnchors [2] allow the user to both passively acquire sensor data via smartphone scanning, and then actively adjust the host device's parameters such as on/off states, via a subsequent wireless connection. These advancements demonstrate the potential of AR to not only augment our perception of the world, but also enhance our ability to shape and modify the world.

However, further development of bidirectional interaction systems faces non-negligible hurdles due to asymmetric hardware capabilities, with passive perception outpacing active manipulation. This imbalance makes active AR less fluent and natural than passive AR, primarily due to two issues: *accurate identification* and *low-latency communication.* Accurate identification relies on scanning and matching identifiers (e.g., QR Codes or visual features) captured by cameras and compared against a local database. However, this process struggles with blurry targets caused by distance, size, or poor lighting, forcing users to reposition. Additionally, matching becomes difficult when multiple nearby targets have similar appearances, compromising both accuracy and speed. Low-latency communication is crucial for smooth interaction by minimizing delays in data transfer between users and target devices. Current systems depend on radio frequency (RF) methods (e.g., BLE, Wi-Fi), which require pre-connection (e.g., pairing or network setup), introducing delays. This is especially problematic in multi-target systems, where frequent device switching reduces efficiency. For example, in human-robot collaboration [12, 34, 38] this delay significantly reduces efficiency. Including all devices in a single network also limits mobility, hindering dynamic applications like multi-drone control. In summary, current bidirectional systems struggle to consistently provide accurate identification and low-latency communication across different scenarios, limiting their scalability and real-world application.

To enable efficient bidirectional interactions with symmetric passive and active capabilities, we propose EchoSight for natural interaction: users simply look at a target for one-step identification and directly communicate with it via an in-situ data link, bypassing the need for separate identification and connection processes. In EchoSight, the "in-situ" data link is activated only when needed, and encapsulates all interaction data locally within the optical link, eliminating the need for external networks. This approach enables flexible interactions in environments without pre-existing network infrastructure or device pre-registration, supporting opportunistic offline engagements. By eliminating additional selection and pairing steps, this look-and-control interaction workflow lets users focus on their tasks, making the interaction seamless and efficient. EchoSight consists of two components: the SightReader (for commercial AR glasses) and the EchoTag (for smart objects). Together, they form a local optical communication channel that enables bidirectional data transfer with high directionality in an echo-like manner, aligning the user's focus with the target object. Specifically addressing the identification challenge, EchoSight allows users to link to a target by simply looking at it. This design reduces the typical two-step identification process to a one-step retrieval, enabling accurate identification with minimal time cost. For the communication challenge, we devise a dual-element optical design that leverages the inherent directionality of optical signals to create an echo-like data channel. This allows for low-latency bidirectional data transmission with minimal interference with out-of-focus targets.

We validate the communication performance of EchoSight with a system evaluation, where metrics such as communication and angle under various lighting conditions are measured. The results reveal that EchoSight supports communication distances up to 5 m and a viewing angle of up to 120 degrees, suitable for most room-scale bidirectional interactions. Additionally, to assess the effectiveness of EchoSight in improving the fluency of Human-Computer Interaction (HCI) tasks, we conduct a user study involving 12 participants, where EchoSight is compared to traditional QR Code scan or voice recognition systems in an emulated Internet of things (IoT) scenario. The results show that EchoSight is able to achieve at least 5x interaction latency reduction, by minimizing the time overhead for identification and connection. NASA-TLX results show lower mental load and improved fluency compared to other methods. Finally, we demonstrate EchoSight through three applications, including industrial robotics, everyday IoT, and AR social media applications.

Contributions.

- We present the design and implementation of EchoSight, a system for bidirectional interaction between AR glasses and smart devices that leverages the high directionality of optical signals to enable an efficient look-and-control interaction scheme.

- We conduct a system evaluation that measures EchoSight's communication performance under various conditions. The results confirm that its hardware and optical design support common room-scale bidirectional interaction applications.
- We conduct a user study that compares EchoSight to traditional scan or voice-based bidirectional interaction systems. By analyzing multi-stage time consumption of interaction tasks in an emulated IoT scenario, we verify that EchoSight improves both interaction fluency and overall user experience.

2 Related Work

2.1 Bidirectional Virtual-Physical Interaction

Classic uni-directional AR systems overlay virtual data onto the real world, allowing passive perception of augmented information, which acts like an expanded smartphone screen [19, 60]. While this provides more space for information, the growing hardware sensing capabilities demand a deeper integration of virtual and physical worlds [71]. Modern spatial computing systems go beyond passive display, requiring tight bidirectional synchronization between virtual data and physical objects [45], with virtual data reflecting real-world object states. Recent studies focus on cross-device synchronization. BISHARE [75] enables bidirectional interaction between a smartphone and AR head-mounted display (HMD) via Wi-Fi, with the HMD tracking the smartphone's physical state (e.g., position, and orientation) and the smartphone controlling virtual objects in the HMD. STREAM [26] achieves similar synchronization between a tablet and HMD, using the tablet as a spatial anchor to control the menu based on the user's gaze. Other studies focus on bidirectional interaction with physical smart objects. eyemR-Vis [28] synchronizes gaze cues between users for AR collaboration over the Internet, using physical markers as relays. HoloBots [27] extends this by enabling bidirectional interaction between a remote user and a local robot, enhancing remote collaboration. These systems use a bidirectional virtual-physical link to improve responsiveness and fluency across fields like education [43, 74], industry [13, 21], and VR entertainment [33, 36].

Recent advances in spatial computing demand not only passive world sensing but also actively altering the physical world. New systems focus on enabling direct manipulation of physical objects, alongside bidirectional synchronization of physical information. For example, MechARspace [76] synchronizes virtual data with physical toys, using Wi-Fi and MQTT to reflect the toys' states in the user interface (UI), allowing users to trigger actions in the toys. Sketched Reality [29] enables virtual sketches to control physical robots via Wi-Fi, creating a shared interaction space. Closest to our work, InfoLED [68] and LightAnchors [2] explore bidirectional interaction by allowing smartphones to acquire information from and control smart devices with LEDs over BLE or Wi-Fi. However, these systems often focus more on passive augmentation, with limited capacity to actively alter the physical world due to slow identification and connection times. Newer systems, like IRIS [31], use contextual images and inertial sensors for point-and-control interaction very similar to EchoSight, but still rely on local network connectivity and pre-registered devices, limiting scalability.

In contrast, EchoSight enhances bidirectional interaction through a novel optical communication design. By using directional optical beams, EchoSight enables instantaneous recognition of targets and seamless interaction without the overhead of identification and connection. All data is exchanged locally, without network dependency. This reduces latency and improves the synchronization of virtual-physical information, thus enhancing the user's ability to actively alter the environment.

2.2 Context-aware Interaction in Smart Environment

Smart environments aim to enhance HCI by utilizing dispersed devices and contextual information, such as target location and user orientation, to optimize workflows and provide customized feedback [14, 69]. Early systems like Smart Sinks [7] automated tasks such as sink height adjustment and used LED feedback for temperature indication. Later research [35, 57] expanded contextual interactivity with gesture recognition [47, 70], optical target identification [59, 67], and tangible displays [40, 66]. With the rise of augmented reality (AR), there is a growing need for faster and more precise contextual data acquisition, including digital IDs. Marker-based methods, such as ARTag [55, 56], attach identity data to objects for accurate localization. However, these methods suffer from limited data capacity, tracking issues in complex environments, and visual intrusiveness. New approaches like InfraredTags [18] and BrightMarker [16] integrate infrared and fluorescent materials into 3D printed objects, improving tracking and reducing obtrusiveness. Other efforts focus on enhancing marker aesthetics and fabrication for broader deployment [16, 22, 64]. Marker-less methods are also advancing. Early techniques, such as SIFT for smartphone tracking [53, 54], have evolved to balance speed and accuracy. Systems like InfoLED [68] and LightAnchors [2] use LEDs in smart devices for optical communication of digital IDs. XR-Objects [17] leverages Computer Vision (CV), SLAM, and large language models for contextual interactions, while BLEselect [73], BLEARVIS [48], and X-AR [8] use RF link capabilities such as Angle-of-Arrival (AoA) and Synthetic Aperture Radar (SAR) for distinguishing objects or penetrating obstacles. A summary of key studies in context-aware and bidirectional interaction is provided in Tab. 1.

Building on previous efforts, EchoSight introduces a novel method for extracting contextual data using optical beams' directionality. Unlike marker-based systems, EchoSight directly extracts contextual data from targets and exchanges data via an ad-hoc optical link, without needing network connections (e.g., Wi-Fi, cellular). This link is temporary, focused on the target, and allows rapid, local exchange of contextual UI and interaction commands, bypassing the time-consuming pre-registration process.

2.3 Optical Wireless Communication for AR/VR

Optical Wireless Communication (OWC) is an emerging wireless technology offering high-bandwidth, highly-directional communication, making it ideal for AR/VR applications requiring seamless, high-speed data transfer. Compared to conventional RF-based technologies, OWC is crucial for immersive AR/VR experiences.

Traditionally, OWC used IR LEDs with simple modulation schemes (e.g., RC-5 [44]) for consumer control, like TV remotes, but lacked angular specificity and user input. Recent research has explored

Table 1: Comparison of feature of current context-aware and bidirectional interaction systems.

Phase		Identification						Communication		
Type		Marker-based		Marker-less				Connection-based		Connection-free
Method		Printed marker	Embedded marker	Visual feature	RF identity	LED pattern	Optical tethering	Local network	RF	Optical tethering
Work	InfraredTags [18] BrightMarker [16]		✓					✓		
	InfoLED [68] LightAnchors [2]					✓		✓		
	SeedMarkers [22] eyemR-Vis [28]	✓						✓		
	XR-Objects [17] BLEARVIS [48] BISHARE [75] STREAM [26] HoloBots [27] MechARspace [76] Sketched Reality [29]			✓				✓		
	BLEselect [73]				✓				✓	
	X-AR [8] IRIS [31]			✓	✓			✓	✓	
	EchoSight						✓			✓

OWC for directional data exchange in AR/VR. Active OWC systems with high-speed transmitters can achieve Gbps data rates for point-to-point links, supporting high-bandwidth AR/VR streaming. For example, Cyclops [23] uses Gbps optical fiber transceivers with free-space optical tracking for VR headsets. Lasertag [11] enhances high-speed target tracking with retroreflective markers for high signal-to-noise ratio (SNR), enabling speeds up to 6.5 m/s. However, these systems require bulky hardware, limiting their use in lightweight, mobile AR scenarios. Passive OWC, using reflective materials for optical tethering, offers a lightweight solution. Early studies demonstrated that high directionality of backscattered light aids in rapid target acquisition [37], though they lacked bidirectional communication. Recent developments added optical modulators to retroreflective materials, enabling bidirectional communication [65], with improvements in range [58] and data rates [63].

While OWC has been explored for high data rates and low latency networking, its application in Human-Computer Interaction (HCI) scenarios, focusing on interaction modality and user attention, has been limited. Building on the latest passive OWC technologies, EchoSight is the first to adapt it for bidirectional virtual-physical interaction in HCI tasks. EchoSight combines OWC's high directionality with gesture control in lightweight AR glasses. It uses a dual-element design with two optical channels for precise alignment, leveraging user attention to bypass time-consuming identification and pre-connection steps. This enables intuitive, look-and-control bidirectional interaction. Since all data is exchanged locally, EchoSight does not rely on external networks, offering flexibility for spontaneous interactions. As a result, EchoSight enhances the scalability and flexibility of future AR applications, providing a streamlined, holistic user experience.

3 System Overview

3.1 EchoSight's Workflow

With EchoSight, when a user intends to interact with a target object, he only needs to look at the target and make the 'point' gesture. As shown in Fig. 2, this will instantly trigger a one-step identification process. The target device ID as well as the corresponding control menu data is directly retrieved from the target device's EchoTag, and this data is parsed at SightReader and then displayed in the UI as a command menu, listing all possible actions and their corresponding gestures. Interactions proceed directly via the infrared link, and once the user's task concludes, simply looking away from the object causes EchoSight to automatically recognize this intention and hide the menu.

Note that with EchoSight, the target is selected directly from the user's focal area, bypassing the need for a selecting menu. Also, all the data transmission is confined to a highly focused optical link, eliminating the prerequisite of pairing or establishing a connection. This creates a look-and-control interaction style that mirrors the natural way humans interact with objects in proximity. This is also the main contribution of EchoSight— by speeding up the target identification without manual selection, and removing unnecessary time overhead caused by the wireless connection, EchoSight streamlines the whole interaction to be more fluent and natural.

3.2 EchoSight's Optical Tethering Mechanism

Similar to common wireless networking techniques like BLE or Wi-Fi, optical tethering establishes a bidirectional data transmission link between two devices. The difference is that optical tethering utilizes optical beams as the data carrier rather than electromagnetic

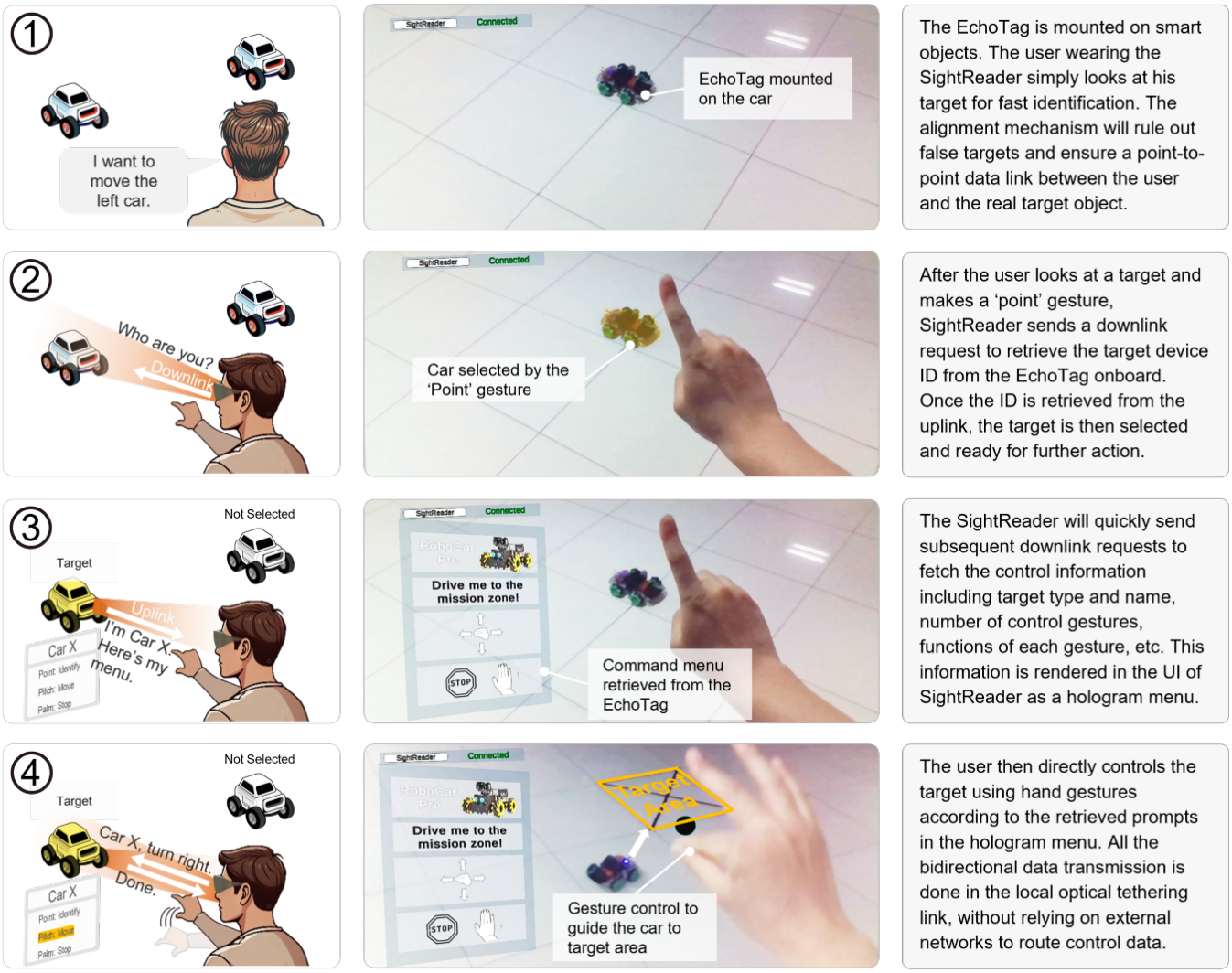


Figure 2: Workflow of EchoSight for bidirectional interaction.

waves as BLE or Wi-Fi. Different from electromagnetic waves that travel freely in all directions, optical beams only travel along a single direction, thus allowing a very concentrated and focused point-to-point communication. In this way, data can be synchronized among two communication devices with minor interference from others, hence the name "tethering" as if two devices are closely tethered together as a whole system.

Among ways of implementing optical tethering, EchoSight utilizes near-infrared backscatter communication [10, 52, 63]. Fig. 3 presents a typical setup for backscatter-based tethering systems. The setup involves two communication devices, namely transceivers. Each transceiver is able to do two actions: sending data to others (called TX), and receiving data from others (called RX). Note that although they can both send and receive data, the underlying functional hardware units are slightly different, as most communication is not symmetric and one of the transceivers acts more proactively while the other more passively. Suppose transceiver A is the proactive one. When sending data to B, the A transceiver embeds data onto the light beam emitted from the LED by modifying the intensity of frequency of the beams, using signal modulation techniques such as on-off keying (OOK) [15] or amplitude-shift keying (ASK) [61]. These modulated light beams are then received by the photodiode on B, and data is extracted at B according to the modulation pattern. This data link is called a downlink. Conversely, when B wants to respond to A, it also needs to embed data onto optical beams. However, as B is the passive one without active units like LED, it utilizes a weaker but more cost-effective sending unit — optical modulator. Instead of directly emitting light beams, B reflects the incoming light beams from A via some reflective material on the surface. Simultaneously, B modifies the states of the optical modulator, usually in the form of liquid crystal spatial light modulators (SLM) or digital micro-mirror devices (DMD) [39]. These

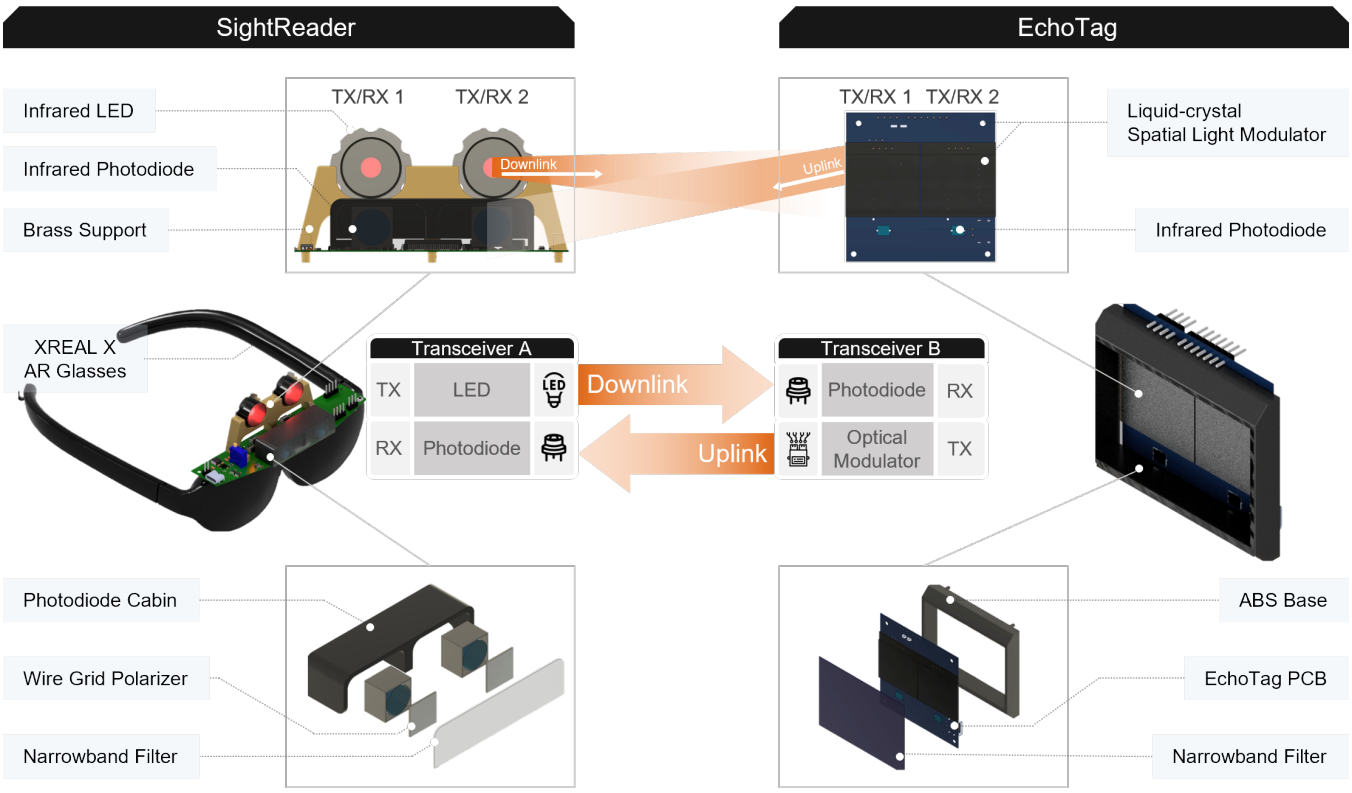


Figure 3: Structure of a typical backscatter-based optical tethering system and structure of EchoSight. A typical bidirectional optical tethering system includes two transceivers A and B. Both of them have one pair of TX/RX hardware units. A uses LEDs as TX and photodiodes as RX. B also uses photodiodes as RX, however, for TX it does not directly emit light but instead reflects the light from A and uses an optical modulator to embed data on the reflected beams. EchoSight improves traditional optical tethering systems with a new dual-element design. In EchoSight, the two transceivers, namely SightReader and EchoTag, each have two pairs of TX/RX, creating two orthogonal infrared data channels. EchoSight leverages the complementary information in these two channels for alignment detection and interference mitigation.

modulators work like camera shutters and can modify the intensity or frequency of the backscattered light beams, thus sending data back to the receiving photodiode on A. This data link is called an uplink. The downlink beams are highly directional due to their optical intrinsic, while the uplink beams are also highly focused since it is backscattered. Synchronizing data within these two links, backscatter-based tethering systems create a highly concentrated local bidirectional data channel and allow for tight data tethering with minor interference.

EchoSight renovates traditional optical tethering systems to cater to the need for symmetric bidirectional interaction scenarios. In EchoSight, transceiver A is called SightReader, and B is called an EchoTag, both of them feature a custom PCB board that can be interfaced with AR glasses or smart appliances. As shown in Fig. 3, EchoSight implements three key designs to enable intuitive interaction. First, we design a dual-element optical front-end. Being "dual-element" means using two groups of TX/RX pairs for the SightReader and EchoTag respectively. By leveraging complementary signals from both groups, we enable alignment detection, preventing false triggers from out-of-focus EchoTags and ensuring rapid, precise identification. Second, we employ two near-infrared bands (850 nm and 940 nm) to create orthogonal, virtually homogeneous optical links. Combined with pulse position modulation [42], this dual-band signal reduces ambient noise, maximizes transmission range and speed, and supports low-latency, connection-free communication for seamless interactions. Finally, EchoTags are responsive—they only activate when the SightReader signals them and immediately fall asleep after communication, ensuring minimal interference from nearby devices and echo-like data transfer.

4 Hardware Design

4.1 Design Goals

We address two key challenges in EchoSight to make the bidirectional interaction more symmetric in both passive perception and active manipulation.

Accurate identification. The first challenge is achieving precise identification with minimal time cost. Traditional systems rely on a two-step process: capturing object features (QR Codes or visual capture) and matching them in a database. While accurate, it is inefficient, creating a speed-accuracy trade-off. This raised

the question: *Is the two-step process necessary?* RF signals from BLE and Wi-Fi are omnidirectional, and ill-suited for directional interactions, while humans naturally engage visually. We concluded that *identification can be simplified to a single step* if the target's identity is directly available. In §4.3, we show how optical signals enable streamlined, one-step interactions.

Low-latency communication. Traditional methods require time-consuming pairing via BLE or Wi-Fi, which introduces delays incompatible with fluid interactions. This led us to ask: *Is connection really needed?* Human communication is *direct and responsive, with no intermediary steps* when focused on a target. Inspired by this, §4.4 discusses our use of infrared backscatter for low-latency, connection-free communication, mirroring natural interactions.

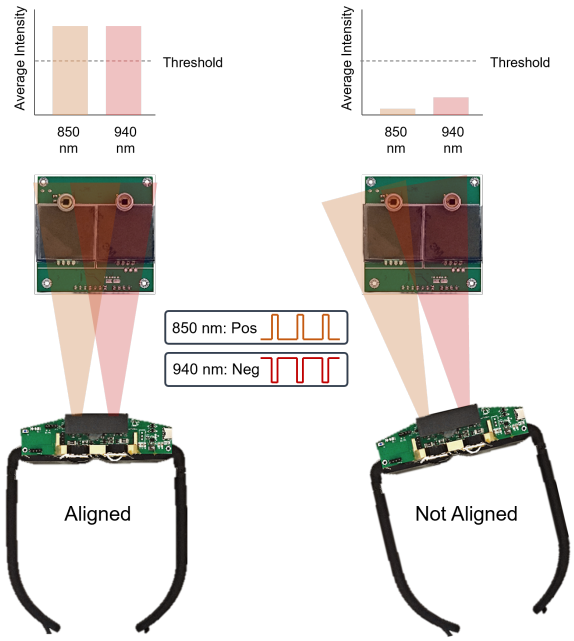


Figure 4: Mechanism of one-step identification and connection-free communication. EchoSight uses a dual-element design for alignment. When SightReader is well aligned with EchoTag, the average DC light intensity detected by two PDs on EchoTag will be both above the predefined threshold and roughly the same. When not aligned, the average intensity of two PDs are not equal with a large difference.

4.2 Hardware Components for Fabrication

We choose the Samsung Galaxy S21 smartphone and XREAL X AR glasses to deploy EchoSight. The system primarily consists of two custom-designed PCB boards, namely SightReader and EchoTag. For the SightReader, we use dual CREE LED (850 nm 1W and 940 nm 1W) for the transmitting, and dual LightSensing LSSPD-2.5 photodiode (PD) as the receiver. Each LED is equipped with a lens with a field of view (FoV) of 5° to converge the light. The board has an area of 7.1 cm x 12.2 cm. The LED supporter is self-designed using Autodesk Fusion 360 and manufactured by CNC machining using H-59 brass material. The receiving cabin shell is designed and 3D printed using PLA material (BASF Ultrafuse). For the EchoTag, the PDs share the same type with SightReader, plus a self-manufactured liquid crystal SLM with an area an 3 cm x 3 cm. The receiving cabin shell is also 3D printed using PLA material (BASF Ultrafuse). The EchoTag board has an area of 6 cm x 6 cm. The whole hardware suite is shown in Fig. 5.

4.3 One-step Identification

To achieve one-step identification and eliminate the need for scanning, it is essential to perfectly align the user's focus with the target object, enabling the object's identity to be instantly relayed back to the SightReader. EchoSight accomplishes this with a dual-element optical design and a complementary signal approach.

EchoSight's design ensures precise alignment using dual-LED transmitters and dual-PD receivers. Illustrated in Fig. 4, each LED and PD on the same side pair up as a transmitting (TX) and receiving (RX) duo, operating at distinct optical wavelengths—850 nm for one pair and 940 nm for the other. Narrow-band light filters eliminate interference from other wavelengths, while narrow optical lenses focus the beams, limiting their FoV. This setup balances the infrared channels' strength as perceived by the EchoTag, and equal distances between LEDs and PDs align the optical center of the beams with the PD receiver's core. EchoSight relies on this phenomenon to rule out false triggering and only wake up the aligned EchoTag, whose identify is then fetched and completes the identification. The identity is then used as an authentication key for subsequent data transmissions. Other EchoTags out of focus will either never receive requests from SightReader due to the high directionality of beams, or will remain silent as long as they think they are not the target, or the identity data does not match. Through this way, EchoSight achieves bidirectional transmission without relying on a connection.

For dependable identity capture, EchoSight employs a complementary signal strategy. Downlink signals of different wavelengths modulate in opposite amplitudes; when the 850 nm channel is high, the 940 nm channel is low, and vice versa. This complementary pattern, named differential signal [9, 51], mitigates sudden environmental optical noise by allowing interference cancellation through signal subtraction at the EchoTag.

An EchoTag discerns its status as the intended downlink recipient by confirming two criteria: a) the received signals' DC components are equal, and b) at any given moment, the voltage levels are opposite, though the decoded data matches. A minimal threshold compensates for slight channel discrepancies due to component manufacturing and assembly variances. In scenarios where identity information is unnecessary—like interacting with a lone object in a room where the user is already familiar with all the commands—this process can be simplified to zero steps for smoother experience.

4.4 Connection-free Communication

To facilitate natural and responsive communication, it is essential to eliminate the need for association without compromising the reliability of the point-to-point data link. We designed EchoSight

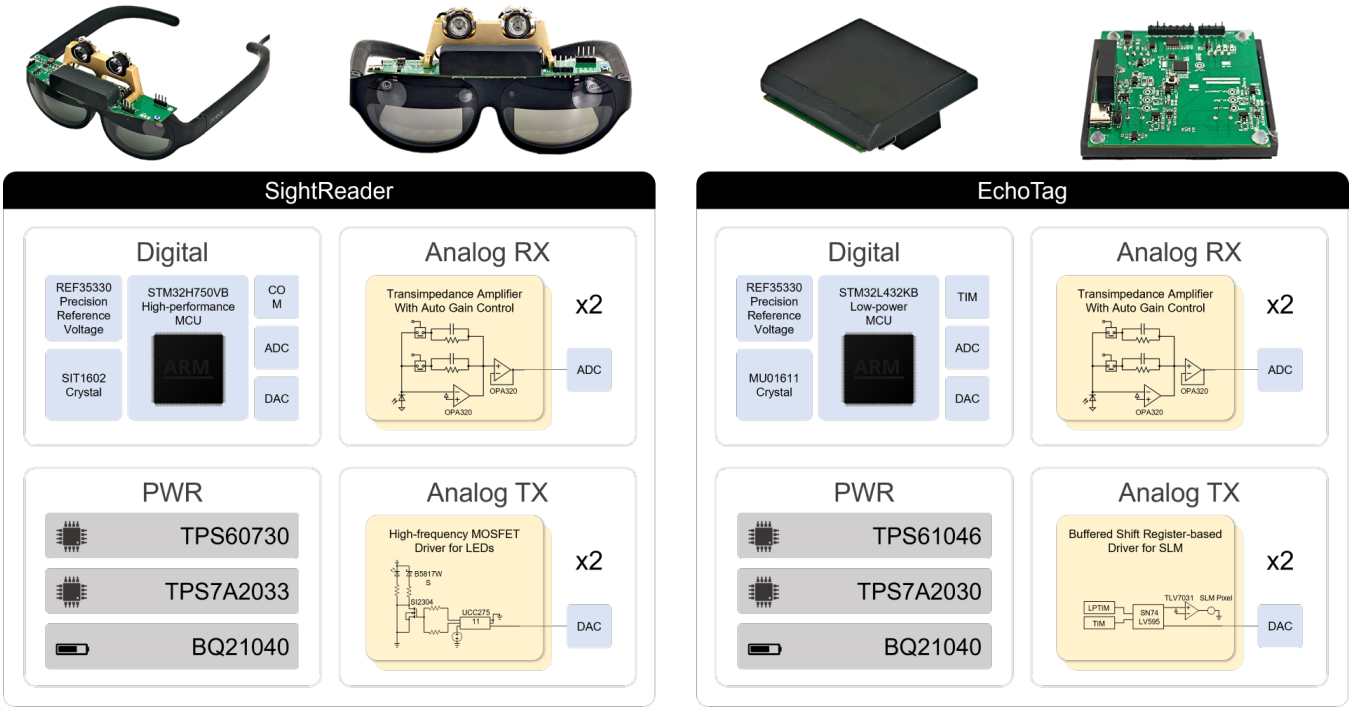


Figure 5: EchoSight hardware suite. For SightReader, a high-performance STM32H750VB MCU is used as the processor to better support ADC/DAC and encoding tasks. A photoconductive trans-impedance amplifier is used to capture infrared beams and convert them to electric signals. This trans-impedance amplifier circuit also features an auto gain control design using analog switches, to cater to different levels of light intensities at different distances. For TX, a fast charge fast sink MOSFET driving circuit is designed to allow fast switching of the LEDs on and off, thus creating modulation. For EchoTag, A low-power STM32L432KB MCU is used. The RX shares a similar design to that of SightReader, with minor additional LC filters. For TX, An SN74LV595 shift registered is combined with the TIM and LPTIM module to cache the driving data for the SLM pixels. The data is then output to the TLV7031 comparator, whose output directly drives the SLM modulator.

to achieve this responsiveness via infrared backscatter communication, which eliminates association requirements by employing an echo-based responsive data link.

Upon receiving downlink data, an EchoTag determines whether it is the intended target using the methods outlined in §4.3. Once confirmation is made, it responds by reflecting the incoming carrier signal to the SightReader. The modulation pattern, including amplitude and frequency changes, is adjusted by dual SLM modulators onboard the EchoTag, one for each infrared channel. The inherent directionality of light beams ensures a perfect backscatter without leakage in other directions, concentrating the energy at the beam's center. Once the initial handshake is done and the identity is returned to the SightReader, subsequent packets will contain the target identity as a key to rule out unwanted targets, which further reduces interference from other EchoTags.

To enhance reliability, data is encoded before being modulated. Optical wireless channels require a high SNR to distinguish the signal of interest from background noise. EchoSight employs pulse-width encoding, transforming raw binary bits into a series of digital pulses of varying widths. This technique transmits a strong, short pulse when data is present and remains idle otherwise, much like spotting a bright blinking star against the dark sky. This method significantly boosts the contrast between the signal and noise, resulting in a higher SNR and thus enhancing the reliability of the data transmission.

4.5 Other Factors

Form factor. In designing the SightReader for EchoSight, we navigated the delicate balance between maximizing communication performance and preserving the glasses' original lightweight mobility. This balancing act imposed significant form factor constraints. Early prototypes ranged from side-hanging boards on the glasses' arms to earphone-like designs dangling from the bottom. Ultimately, we settled on a crab-like design with the SightReader mounted atop the glasses. This design, as depicted in Fig. 5, synergizes with the glasses' top platform, serving both as a holder and effectively distributing most of the weight onto the nose pad, thus enhancing the wearability of the device.

Power and heat control. Power consumption and heat dissipation were also critical considerations. The dual LEDs on the SightReader, essential for transmitting data, raised concerns about power usage and potential overheating, which could endanger the

user's skin. We tackled these issues through two strategies. Firstly, to minimize power consumption, we employed pulse-width encoding, allowing the LEDs to remain idle for extended periods and only activate briefly for pulses. This approach significantly cuts down the board's power usage compared to continuous LED operation. The incorporation of low-power DCDC chips and operational amplifiers further aids in reducing energy consumption. Secondly, for heat dissipation enhancement, the SightReader utilizes a four-layer PCB design with extensive copper pours to help dissipate heat. Additionally, we developed a CNC-machined LED support using brass, as in Fig. 3, which not only stabilizes the LEDs for easier assembly but also uses brass to efficiently transfer heat away from the device. These design choices have been proven to keep temperature increases under 10 degrees Celsius, a safe range for human skin contact.

5 Software Design

5.1 Android App

We designed an Android app for EchoSight, which runs on the backend computing smartphone. The whole Android project is written in C sharp language in Unity version 2022.3.8 f1, with XREAL's official SDK for XREAL X glasses NRSDK v2.2.0. When building the app, an Android Studio project is generated from Unity and compiled, creating an app file to be installed on the phone. This app is finally run on the phone and displayed in the glasses.

5.2 User Interface

The UI design of EchoSight employs a straightforward interaction logic: when users focus on a target object, pointing and lightly tapping with a finger, a guide menu emerges from the left side of the AR interface after the identification is done. The menu offers what gestures and their functions are supported for further operation. Should the user's attention shift away from the object, the menu vanishes. The elements such as images, buttons, and labels are conveyed in XML format. The UI elements are first serialized by the transmitter into an XML-compliant string for optical transmission, and de-serialized to reconstruct the XML object at the receiver, similar to how a HTML Document Object Model (DOM) is delivered on web pages. This is done by utilizing C sharp's XmlSerializer library. The reconstructed objects are then used to populate NRSDK's predefined UI containers for displaying. Note that real-time contextual data from the target EchoTag is utilized for each interaction, eliminating the need for a matching database or network-based data retrieval, as all exchanges are local and on-demand. SightReader identifies EchoTag and triggers interactions using only the locally exchanged data, offering scalability and flexibility for walk-in scenarios such as museum tours, urban navigation, and interactive learning.

5.3 Indicator and Feedback

A circle crosshair aiming indicator is drawn in the center of the UI to help the user better aim for the target device when doing one-step identification. To help users get a grasp of when the identification is triggered and done, we add acoustic feedback to the system. When identification is done and both device ID and menu content are fetched, the system will play a 'menu-open' sound effect. When the user looks away from the previous target device, the system will recognize this focus shift and automatically close the previous hologram menu, with a "menu close" sound.

6 System Evaluation

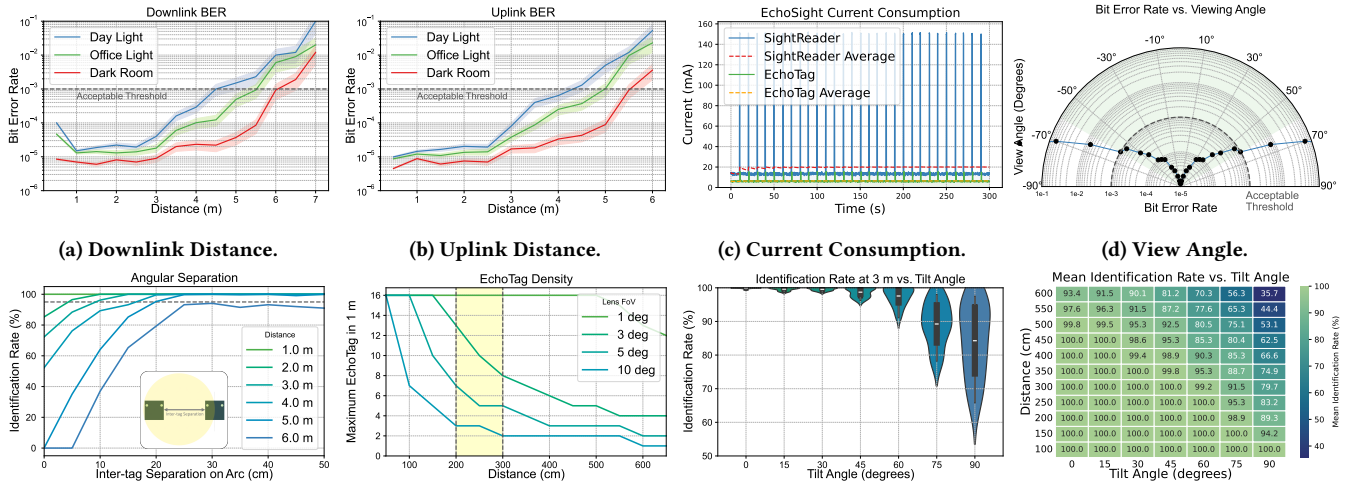
In system evaluation, we focus on evaluating pure hardware performance, to explore whether the optical link itself is robust and capable of supporting common room-scale AR/VR applications. Two aspects are considered: 1) the communication performance, including the maximum communication distance, view angle, and data rate. 2) power-related metrics, including the power consumption and dissipation.

6.1 Setup and Procedure

Setup. Our system evaluation utilized XREAL X glasses paired with a Samsung Galaxy S21 smartphone, which served both as the power source and the computing platform for the glasses. This setup choice by XREAL aimed to address the glasses' limitations in handling compute-intensive tasks independently. We developed the data processing pipeline in Kotlin for Android and deployed the app on the smartphone. This app interfaces with the AR wrapper app Nebula to automatically enter AR mode and present a holographic UI upon activation.

Procedure. For assessing communication capabilities, we conducted experiments under three lighting conditions: direct sunlight (1000 lux), standard office lighting (400 lux), and darkness (100 lux). We tested downlink and uplink performance across various distances and angles, recording the overall data error rate to gauge reliability. Furthermore, we determined the maximum data rates for both uplink and downlink communications at a 3-m distance under typical office lighting conditions. This was achieved by continuously transmitting a predefined 32-byte message and monitoring the error bits and instantaneous bit rate. All experiments were repeated thrice, with results reported as averages. For power metrics, we employed a Monsoon HV Power Monitor to track the voltage and current consumption of both the SightReader and the EchoTag. Additionally, we used a FLIR Lepton 3.5 thermal camera to observe the SightReader's power dissipation. Power consumption was measured while the SightReader transmitted random data every 10 seconds, simulating a standard interaction pattern.

To evaluate the impact of angular specificity on EchoSight's real-world usability and scalability, we conducted two experiments. First, we measured the minimal conflict-free distance between two EchoTags. With the SightReader fixed, we placed the EchoTags on an arc equidistant from the user. The left EchoTag was aligned with the SightReader. We gradually reduced the angular separation between the EchoTags and used the SightReader to identify the left EchoTag at each position. A 32-byte random string was sent to and returned from the EchoTag, with identification considered successful if the returned string matched the sent one. We repeated the process 1000 times per position and calculated the conflict-free identification rate as the ratio of successful identifications to total transmissions. A 5° lens was used for the SightReader. Next, we determined the maximum EchoTag density without performance degradation. We repeated the previous experiment with different lens FoV settings (1°, 3°, 5°, and 10°). For each FoV, we identified the



(a) Downlink Distance. (b) Uplink Distance. (c) Current Consumption. (d) View Angle.

(e) Angular Specificity. (f) Density. (g) Tilt Angle Single. (h) Tilt Angle Multiple.

Figure 6: Results for the system evaluations. (a)(b) The curves below the acceptable threshold represent the working ranges under each condition. (c) The dashed lines represent the averaged current for SightReader and EchoTag over time. (d) This is a polar graph representing the BER with respect to different view angles. From origin point, each half circle represents a same BER level. For example, the dashed half circle represents a BER of 10^{-3}. The green shaded area represents the FoV under the acceptable threshold, which is -60° to 60°. (e) The inter-tag separation is depicted in centimeters for intuitiveness, defined as the distance ensuring no interference between two EchoTags. Specifically, it is the separation at which the beam projection area fully covers the left EchoTag without crossing the center line of the right EchoTag. The alignment detection algorithm detailed in Fig. 4 ensures the silencing of the right EchoTag. (f) The curve illustrates the maximum number of deployable EchoTags within a 1 m arc length under various distances and lens FoV settings. The highlighted area represents the common range for most look-and-control interaction scenarios. The upper limit is constrained by the current form factor of EchoTag. (g) The violin plot displays the success identification rate when communicating with a stationary EchoTag as the SightReader tilts. The inner boxes indicate the precise data distribution, while the outer violin shape serves as a concentration indicator. (h) The heatmap represents the mean identification rate at various distances when the SightReader is tilted. Note that extreme tilting significantly affects performance at larger distances.

minimum separation that maintained a 95% identification rate and calculated the maximum number of EchoTags deployable within a 1-m radius without conflict.

Finally, we evaluated the influence of head tilting through two experiments. First, we assessed the tilt effect on a single EchoTag. We positioned an EchoTag 3 m from the SightReader, both mounted on tripods. The SightReader was attached to a servo for precise angular rotation. We rotated the SightReader from 0° to 90° in 15° increments. At each angle, we repeated the identification process 1000 times as in the separation experiment and recorded the results. Next, we evaluated the tilt effect at various distances by moving the EchoTag to different positions and repeating the identification process. To capture real-world performance, for the same distance, different locations with various lighting conditions were mixed, and the average results were recorded.

6.2 Communication Performance Results

We can see from Fig. 6a and Fig. 6b that lighting conditions have a significant impact on communication reliability. If we take 10^{-3} as the threshold the communication is considered reliable, the downlink and uplink of EchoSight can work up to around 5.5 m and 5.0 m respectively. For the viewing angle, we can conclude from Fig. 6d that at a threshold of 10^{-3}, EchoSight supports around $\pm 60^\circ$, in total a 120°. In other words, if the angle between the user's location and the normal line of an EchoTag is within $\pm 60^\circ$, he can communicate with it by looking at it without confusion. This performance allows EchoSight to support common room-scale directional interaction applications.

We measured the maximum supported data transmission rate for the bidirectional communication. The downlink features a raw data rate of up to 110 kbps without any error correction code applied, and the uplink supports up to 512 bps. This asymmetry is limited by the inherent working principles of SLM modulators [4]. SLMs are commonly used in monitor screens. They modulate light by changing the arrangement of internal liquid crystal molecules, limiting their maximum refresh rate to the hundreds of Hz. Since in most responsive interactions, most information is sent in the downlink and the uplink mainly carries responses, we argue that the current data rate is able to support common bidirectional scenarios.

6.3 Power Performance Results

The power performance of EchoSight demonstrates high efficiency, as detailed by the current consumption patterns of both the SightReader and the EchoTag, illustrated in Fig. 6c. The data

shows that the power consumption directly correlates with the pulse emissions. The SightReader exhibits a peak current consumption of approximately 150 mA, yet its average consumption is maintained around 20 mA, showing the effectiveness of pulse-width encoding in minimizing average power use. Meanwhile, the EchoTag, equipped with low-power integrated circuits (ICs) and microcontrollers (MCUs), along with the SLM modulators that draw minimal current, shows an average current consumption under 10 mA. Given a 3 V power supply, the EchoTag consistently consumes tens of milliwatts of power. In real-world deployment, EchoTags are not continuously active; they only activate for on-demand transmission, significantly reducing average power consumption for long-term use. For ultra-low-power applications, the onboard MCU's deep sleep mode and scheduled power cycling can further decrease the average current consumption to microamperes, enabling EchoTags to operate on standard batteries for years.

Regarding power dissipation, the temperature increase within the LED area was measured to be around 7° Celsius. Given room temperature conditions, such a temperature rise does not pose discomfort to human skin. These findings validate the design choices discussed in §4.5, confirming their practicality and efficiency in managing power consumption and dissipation.

6.4 Scalability and Usability Results

Fig. 6e shows the identification rate as a function of inter-tag separation, reported in centimeters of arc length. As angular separation increases, beam coverage avoids out-of-focus tags, improving the identification rate. For an acceptable threshold of 95%, at typical interaction distances (up to 3 m), the minimum distinguishable separation is 10 cm edge-to-edge. This allows for high-density deployment without confusion, as most devices are placed tens or hundreds of centimeters apart. With a fixed 5° lens FoV, the system can distinguish tags as close as 10 cm without performance degradation, such as misidentification or selecting multiple tags. Minor misalignment is tolerated, as long as the signal difference is within the threshold, as described in §4.4.

Fig. 6f shows results with varying lens FoV settings. For the typical 2-3 m interaction distance, a 5° lens supports up to 7 tags within 1 m. For higher-density scenarios, narrower lenses are effective; a 1° lens supports up to 16 tags within 1 m at a 5 m distance. However, due to the tag size (6 cm × 6 cm), a hard limit of 16 tags applies. To support higher densities, narrower lenses, and customized form factors can be used.

Fig. 6g presents communication results with a single EchoTag at various head tilt angles. The identification rate stays high (> 90%) up to 60°. Beyond this, it drops sharply, reaching around 70% at 75°, with increased variance. This is likely due to the SightReader's beams rotating with the head, ensuring consistent intensity at moderate tilts. However, at larger angles, the beams' optical centers deviate from the EchoTag's PDs, increasing susceptibility to noise and uneven backscattering. Overall, the system performs well with head tilts up to 60°, and while extreme tilts are rare in typical interactions, performance decreases beyond this threshold.

Fig. 6h shows the mean identification rate across tilt angles and distances. The rate remains above 90% for up to 30° at all distances, though averaging across various lighting conditions results in a slight decrease compared to Fig. 6g. As tilt angles increase, the rate decreases, especially at longer distances. For instance, at a 90° tilt, the rate drops to around 53% at 500 cm but stays above 80% at shorter distances. While large tilts are uncommon, future applications may require more aggressive error correction and retransmission protocols to support higher tilt angles.

7 User Study

In the user study, we aim to evaluate the usability of EchoSight. Specifically, we want to verify whether the two design ideas of EchoSight, namely one-step identification (§4.3) and connection-free communication (§4.4), can effectively improve the overall interaction experience in daily AR applications, compared to traditional scan-based methods (*e.g.,* QR Code) and voice-based methods (*e.g.,* voice control).

7.1 Setup

As shown in Fig. 7, we set up an emulated IoT environment with three devices: one smart lamp, one RGB night light, and one smart speaker. Three interaction methods are provided for each device: scan-based method using QR Code, voice control method, and EchoSight. As described in Tab. 2, for each method, we split the whole interaction process into multiple phases. The detailed span of each phase is summarized in Fig. 8.

Each device is integrated with a BLE module for wireless connection, a QR Code containing the device ID for the scan method, and an EchoTag to enable interaction with our system. For a fair comparison, the QR Code is printed with an equal size to the EchoTag (*i.e.,* 6 cm x 6 cm). For EchoSight, we use XREAL X AR glasses as the base AR platform, and a SAMSUNG Galaxy S21 as the computing backend.

We recruited 12 participants (7 males, 5 females, mean age = 24.2). Participants have various experiences with VR/AR systems: 3 had extensive experience, 4 had light familiarity, and 5 had little or no experience. Participants were seated on a sofa 2 m away from three devices. They are allowed to stand up and move closer if preferred to better scan or interact with the devices.

7.2 Procedure

The study started with a short introduction and training session. Participants were familiarized with the AR glasses, SightReader, and EchoTag, followed by a 10-minute demonstration of the prototype's operation. This included learning how to activate AR mode on the XREAL glasses, navigate the holographic UI using gaze and gestures, and interact with IoT devices using EchoSight 's features like point-to-identify and pitch for further commands. Following this introduction, participants practiced simple tasks such as adjusting light brightness or hue.

After mastering the basics, participants undertook a high-level task independently. The task includes a timed session and a free session. In the timed session, the users were asked to complete three out of six pre-defined actions (*i.e.,* lamp brightness up/down, RGB light hue warmer/colder and speaker on/off), according to a randomly generated action list. The order of the actions is randomly arranged and counter-balanced, so it may require users to switch between three targets during the session. They were asked

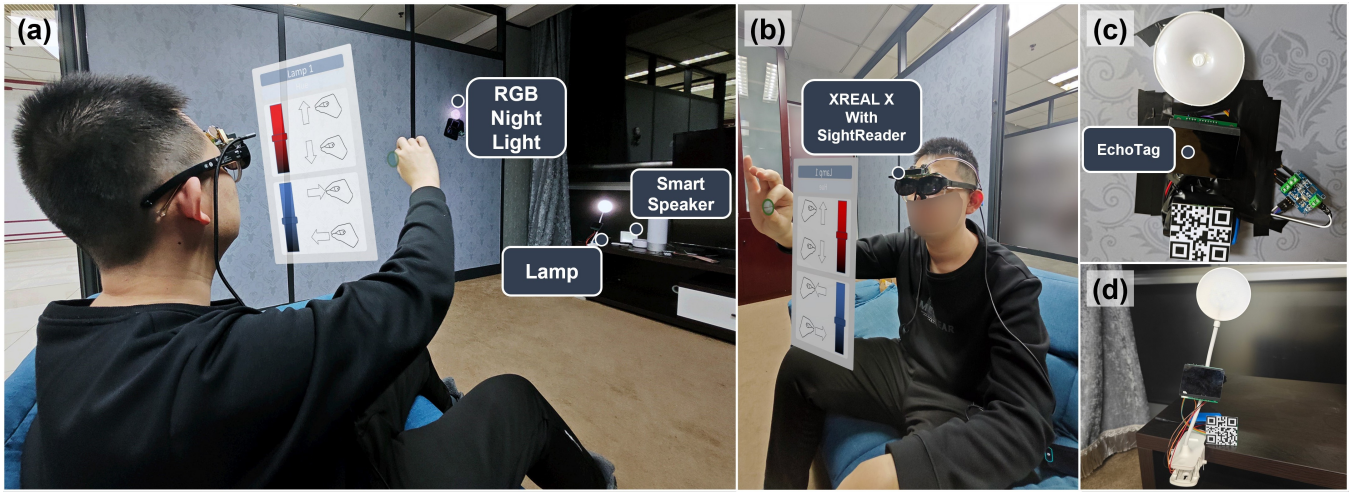


Figure 7: The user is conducting the study in the emulated living room. (a)(b) The user sends hue control commands to the target RGB light according to the prompts displayed in the hologram UI and uses the 'pinch' gesture to change the warmth of the RGB color by moving up/down and left/right. (c)(d) An EchoTag and an equal-sized QR Code are equipped with the RGB light and the smart lamp.

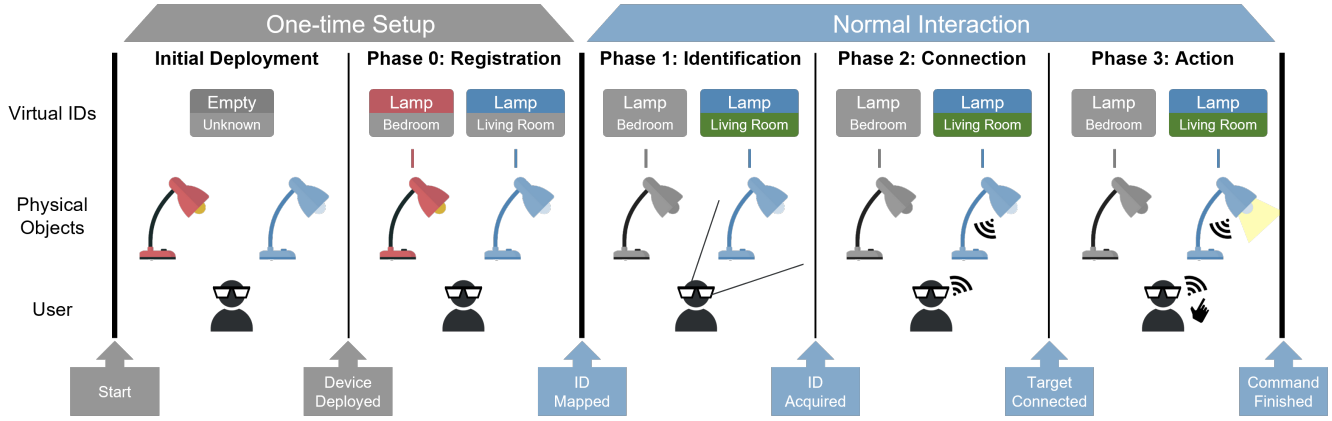


Figure 8: Phases of the whole interaction. Deployment and phase 0: New devices need registration to map their virtual IDs in the database to the physical objects. This registration process is a one-time setup and is unnecessary in the interaction afterward. Phase 1 - 3: These are the decomposed phases for a normal interaction process, recorded in the study.

to complete the actions three times in total, each time using one of the provided interaction methods.

To investigate the interaction details, we recorded the time consumption during each phase of normal interaction and calculated the mean value and the standard deviations for each phase using each method. The span of each phase is described using arrows in phase 1-3 in Fig. 8. Note that registration is a one-time setup before using these IoT devices, and its period is out of the scope of the study and thus not recorded. This multi-phase time data decomposes a normal interaction into finer-grained parts and helps to gain objective insights into the benefits and drawbacks of each method.

In the subsequent free session, participants experimented with all three methods without specific objectives, allowing for improvisation. Afterward, they were asked to complete a NASA-TLX questionnaire addressing mental and physical load, fluency, and frustration on a 1-7 Likert scale, followed by a brief interview to gather additional insights. These steps aim to collect and quantify subjective feedback from the users over each method.

7.3 Study Results

Multi-phase time consumption. Fig. 9 shows the result for the three stages during a normal interaction process. For the identification phase, EchoSight presents a 5x time reduction compared to two other traditional methods. This is reasonable due to the

Table 2: Explanation of Three Interaction Methods.

	Phase 0: Registration	Phase 1: Identification	Phase 2: Connection	Phase 3: Action
Scan	Manually mapping decoded QR Code ID with physical device	Look at the device and hold still to scan	Need BLE or Wi-Fi connection	Gesture control with the UI buttons and sliders
Voice	Manually mapping device name with physical device	Say the device name	Need BLE or Wi-Fi connection	Say the command
EchoSight	No registration	Look at the device	No connection, data is communicated point-to-point	Gesture control according to the prompts in UI

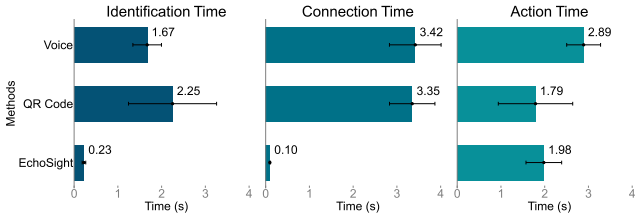


Figure 9: The time consumption for each phase of the task using each method. The mean value is shown next to each bar and the standard deviation is shown as error bars.

one-step identification design. Specifically, by reducing the traditional scan-and-match routine to a simple direct ID acquisition via the optical link, EchoSight effectively removes the extra overhead caused by QR Code scanning or completely says the name of the device. We also observe two interesting facts about the two other methods: the scan-based method is strongly affected by the habits of the user: The randomness of steadiness when scanning, whether choosing to move closer for better scan quality, all bring a large variance to the total identification time. Also, a two-sample t-test and an F-test show that voice command has a small mean value and a smaller variance compared to QR Code scanning (both cases $p < 0.001$). We observe that for voice identification, the time for saying the device name is relatively linearly correlated to the length of the name. The only factor contributing to the variance is a different speak speaking speed. The results for this phase are evident that the one-step identification design in EchoSight is able to largely improve the identification latency by both reducing the mean identification time and bringing a smaller and controllable variance.

For the connection phase, both the QR Code scanning and voice control methods need several seconds to establish a wireless data link. Two-sample t-test and F-test do not find a significant difference between them ($p_t = 0.75$ and $p_F = 0.2$). This is due to the same BLE 5.0 technology used as the underlying wireless protocol. With EchoSight, since there is no connection and the data is directly bidirectionally shared in the optical tethering link, this connection time is significantly reduced to a short period for authentication and point-to-point channel establishment. Note that in practice, for some cases like smart home and smart office, devices are very likely to be pre-registered and maintain a permanent BLE or Wi-Fi connection to a base station or a router. In this case, the latency of this phase for the scanning and voice method can be reduced to the same order at several milliseconds. However, for other applications involving more mobile and random opportunistic interaction between multiple users and multiple target devices, since no pre-registration is possible, the connection would still bring a lot of overhead. For example, interacting with random retailing robots in public airports or museums still requires connection switching across different areas with different BLE or Wi-Fi networks. In this sense, the design of connection-free communication in EchoSight is able to keep a consistent low latency in all cases.

Finally, in the action phase, all three methods present a relatively long time consumption. The voice method is longer since saying out the complete command and completing the recognition requires some time. For QR Code scanning and EchoSight, users interact with the virtual hologram UI elements or directly manipulate them with their hand gestures. Two-sample t-test presents no significant difference between the mean value ($p = 0.31$). Although the contribution of EchoSight does not cover this phase, we do believe that the latency of this phase can be further reduced using new interaction technologies like eye-tracking [20].

Subjective feedback. The questionnaire result shown in Fig. 10 focuses on mental load, physical load, fluency, and frustration to gauge users' subjective feedback across three methods. For all four metrics, both EchoSight and voice achieve a better score. For mental load and fluency, the majority of the scores for voice and EchoSight lie in the range of 1-4. This shows that users prefer interaction schemes that resemble human nature, be it responsive look-and-control like EchoSight, or voice control like using Siri. We can also see that EchoSight gives more sense of fluency to users compared with other methods. Specifically, we notice that the QR Code scan-based method does not perform well in all of the four metrics. To figure out the reason for this, we collect relevant feedback from the final interview. Here's one of them from P7:

> "The problem is not the QR Code itself, but the extra effort I need to move to the right position, keep still for a while, and get it scanned. This lagging is hard to ignore and a bit annoying when you have to switch between targets, and the whole interaction feels like two halves, not as holistic and natural as the other two methods."

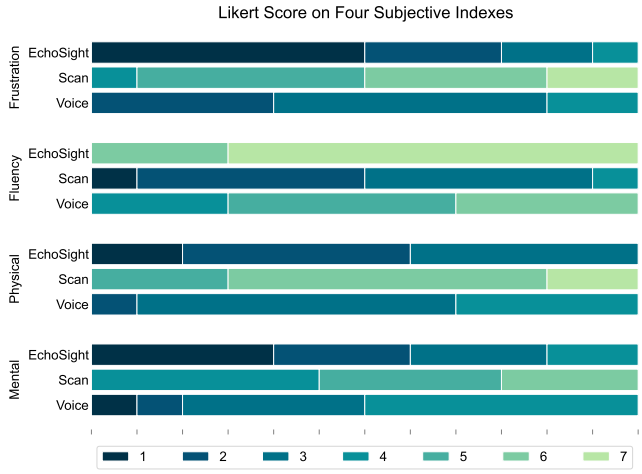


Figure 10: Likert scores for the questionnaire. Each tick represents one score. For each method on each index, a total of 12 scores are collected.

Feedback from other participants also indicates the same underlying idea: one-step interaction feels more natural. With EchoSight and voice control, the user is able to interact with smart objects in a way similar to natural interaction, by simply looking and controlling or voice. With scanning, the process is split into a scan stage and an interacting stage, making a less fluent experience. The participants also mentioned that the voice method is not suitable for privacy-sensitive cases like public services, while EchoSight is still able to keep silent and keep the interaction confined to the user's context.

In conclusion, participants appreciate the design of EchoSight and find it natural and smooth to use. Both the quantitative and the qualitative results affirm the design of EchoSight helps to improve the overall fluency in bidirectional interaction scenarios.

8 Application

We demonstrates three applications with EchoSight's new bidirectional interaction support, shown in Fig. 11.

8.1 Multi-target Human-robot Collaboration

Within the industrial sector, human-robot collaboration remains a critical challenge, marred by non-intuitive interactions and inefficient switching between multiple robots. In terms of interaction processes, EchoSight employs optical methods for fast target identification, allowing for precise interaction without the need for additional switching operations. Regarding interaction logic, once aligned, simple hand gestures can be used to bring up a detailed collaboration menu and access the target robot's status information. The continuous collaboration commands are transmitted entirely in the optical data link via connection-free communication, creating a reliable low-latency performance. This capability has the potential to make collaboration more intuitive and efficient, and speed up the critical production steps of the whole industrial production pipeline.

8.2 IoT Smart Office/Home Control

This application demonstrates the capability of our system to accomplish a variety of target tasks indoors, embodying the vision of the IoT to achieve a connected world. For instance, within an IoT-enabled smart office setting, we can control the printer to switch on and print specific documents, and adjust the brightness of the workspace, all without leaving our seats, enhancing office productivity. In the context of an IoT smart home, lying in bed, we can simply use EchoSight to check the indoor temperature and adjust the air conditioner, or with a simple gesture, turn on the TV right in front of us to watch our favorite program. This look-and-control interaction illustrates the potential of integrating EchoSight with IoT for more efficient interaction, creating a more personalized living and working environment.

8.3 AR Social Application

EchoSight can enhance the social experience in future AR-based social applications. When wearing EchoSight, users can obtain the electronic business cards of their interaction partners simply by looking at them during social activities and can save their information with simple gestures. Users can also display their Facebook updates on the EchoSight interface, where viewers can like posts with a simple thumbs-up gesture. Compared to the ease of losing paper business cards and the inefficiency of scanning QR Codes or searching to add friends on mobile phones, the EchoSight system better aligns with the natural habits of interpersonal communication. By focusing on someone, you can simultaneously retain important information and display comprehensive background details, while remaining silent. This design greatly improves the efficiency and quality of interpersonal interactions.

9 Discussion and Future Work

Security and privacy. EchoSight leverages the directionality of optical beams for alignment, making it challenging for attackers to eavesdrop on the optical channel without being precisely aligned along the straight optical path. This significantly raises the barrier against casual eavesdropping and man-in-the-middle attacks. In future work, we plan to integrate well-established security protocols from web and RF technologies to enhance EchoSight's security. For SightReader side, we will explore biometric authentication [6] to prevent unauthorized access. On EchoTag side, digital signatures [30, 41] can deter malicious fake nodes, and Physical Unclonable Functions (PUFs) [25] can generate unique, unclonable digital identifiers. For the communication channel, we will implement asymmetric encryption, such as ECDH [24], for secure key exchange and symmetric encryption, such as AES [1], for subsequent secure communication. This will thwart rogue transmitters from deciphering intercepted data, even if they manage to intercept it along the direct path. In terms of privacy, compared to previous CV-based work that relied on the camera for scanning, EchoSight only uses PDs for reception, which has no spatial resolution and thus does capture user-specific data. For sensitive application-layer data, hardware encryption technologies, such as Hardware Security Module (HSM) [62] can be employed to securely store data on-board and prevent data leakage. As a future direction, we recommend investigating the roles and responsibilities of system controllers and

Figure 11: Three applications of EchoSight. (a) With EchoSight, a user can control robot cars or drones in sight directly with gestures. The directionality of EchoSight will ensure a point-to-point control between the user and the target robot. Switching between multiple robots no longer requires complex connection switching, but is as simple as looking at them. (b) In daily IoT scenarios, users can interact with different IoT devices scattered around the room, without moving around or selecting them in a menu. Most common functions like selecting documents for printing, and turning on/off fans or lights, can be completed by looking and a simple "pinch" gesture. (c) In AR social applications, the users can look at a stranger and unobtrusively fetch the public information such business card. The feedback of EchoTag will make sure the user gets the wanted information while remaining silent.

users in interactive environments, particularly regarding control over critical components like IR transmitters. A Role-based Access Control (RBAC) system [46] could regulate user access, while real-time monitoring and anomaly detection could prevent rogue signal interference. Additionally, we advocate for the development of industry-wide security standards, including encryption, authentication, and intrusion detection, to ensure safe, secure interactions. These efforts will help establish trust in interactive technologies like EchoSight and promote their widespread adoption.

Relevance to future networking technology. EchoSight offers a novel physical-layer solution that is transparent to higher-level protocols, enabling bidirectional interaction by eliminating manual selection and accelerating communication through a connection-free optical tethering approach. While future networking technologies such as 6G or 7G networks may offer improved data rates and reduced transmission delays, the current bottleneck in bidirectional interaction, as depicted in Fig. 8, is not the networking technology itself but the prerequisite phases, particularly the identification and connection establishment phases. Therefore, EchoSight's capacity to streamline interactions by simplifying device identification and communication setup will complement future connectivity enhancements, positioning it as a versatile solution for next-generation, high-density, and low-latency smart environments. We envision a future where directional technologies like EchoSight are utilized for the initial pre-communication handshake phase, with other networking modalities serving as the follow-up high-throughput, low-latency communication carriers.

More flexible alignment. The implementation of EchoSight currently relies on head orientation for system alignment, which suits a wide range of applications. However, for scenarios demanding ultra-precise alignment, incorporating eye-tracking technology could significantly enhance accuracy. Future iterations of EchoSight could employ Micro-Electro-Mechanical Systems (MEMS) mirror technology to adjust the direction of downlink light beams in real-time, aligning with the user's eye movements. Using head position to determine general focus and eye direction for pinpoint accuracy, this dual-level control promises to refine interaction granularity further. Integrating this feature with advanced AR devices like the Apple Vision Pro, which already boasts mature eye-tracking capabilities, could unlock new, precise application scenarios. For customized, lightweight solutions like EchoSight, this will necessitate a more complex optical front-end, featuring dual galvanometers for 2D beam steering and an adjustable aperture for fine-tuning the beam projection area. Additionally, it will require lightweight tracking and focusing algorithms capable of running on mobile devices. While this advanced customization is beneficial, it is non-trivial and thus we reserve it for future work.

Asymmetric data rate. Additionally, while EchoSight currently supports a downlink data rate of approximately 110 kbps, the uplink data rate is limited to 512 bps. This asymmetry may not impact many current applications but could limit future use cases requiring real-time, high-speed tethering in both directions, potentially reaching several Mbps. To accommodate these demands, transitioning from LED-based communication to laser technology could provide a solution. Lasers would not only enable more precise alignment but also harness the full potential of backscatter-based data links, offering high-speed, low-power laser transceivers for enhanced bidirectional communication. This evolution of EchoSight technology could vastly broaden its applicability, from gaming and industrial

automation to intricate IoT ecosystems and beyond, setting a new standard for AR interaction fidelity and responsiveness.

10 Conclusion

In this paper, we introduce EchoSight, a system for AR glasses that provides efficient bidirectional interaction. Utilizing infrared backscatter communication, EchoSight addresses two key challenges of traditional QR Code scanning and voice control methods, namely accurate identification and low-latency communication. By reducing identification to a one-step process and offering flexible, connection-free communication, EchoSight enhances AR's capability to actively alter the physical world, while providing a holistic bidirectional interaction. Through system evaluation and a user study, we demonstrate that EchoSight improves interaction efficiency and fluency in common bidirectional AR scenarios.

Acknowledgments

We thank the anonymous reviewers for their insightful comments. This work is supported by National Key Research and Development Plan, China (Grant No.2023YFB2903902) and National Natural Science Foundation of China (Grant No.62272010 and 62061146001). Chenren Xu (chenren@pku.edu.cn) is the corresponding author. Jingyu Li, Qingwen Yang and Kenuo Xu are affiliated with School of Computer Science at Peking University. Chenren Xu is affiliated with School of Computer Science at Peking University and Key Laboratory of High Confidence Software Technologies (Peking University), Ministry of Education.

References

[1] Ako Muhamad Abdullah et al. 2017. Advanced encryption standard (AES) algorithm to encrypt and decrypt data. *Cryptography and Network Security* 16, 1 (2017), 11.

[2] Karan Ahuja, Sujeath Pareddy, Robert Xiao, Mayank Goel, and Chris Harrison. 2019. LightAnchors: Appropriating Point Lights for Spatially-Anchored Augmented Reality Interfaces. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 189–196. doi:10.1145/3332165.3347884

[3] Syed Shah Alam, Samiha Susmit, Chieh-Yu Lin, Mohammad Masukujjaman, and Yi-Hui Ho. 2021. Factors Affecting Augmented Reality Adoption in the Retail Industry. *Journal of Open Innovation: Technology, Market, and Complexity* 7, 2 (2021), 142. doi:10.3390/joitmc7020142

[4] Sahar Ammar, Osama Amin, Mohamed-Slim Alouini, and Basem Shihada. 2022. Design and Analysis of LCD-Based Modulator for Passive Sunlight Communications. *IEEE Photonics Journal* 14, 5 (2022), 1–17. doi:10.1109/JPHOT.2022.3200833

[5] Hadi Ardiny and Esmaeel Khanmirza. 2018. The Role of AR and VR Technologies in Education Developments: Opportunities and Challenges. In *2018 6th RSI International Conference on Robotics and Mechatronics (IcRoM)*. IEEE, Tehran, Iran, 482–487. doi:10.1109/ICRoM.2018.8657615

[6] Debnath Bhattacharyya, Rahul Ranjan, Farkhod Alisherov, Minkyu Choi, et al. 2009. Biometric authentication: A review. *International Journal of u-and e-Service, Science and Technology* 2, 3 (2009), 13–28.

[7] Leonardo Bonanni, Ernesto Arroyo, Chia-Hsun Lee, and Ted Selker. 2005. Smart sinks: real-world opportunities for context-aware interaction. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems* (Portland, OR, USA) *(CHI EA '05)*. Association for Computing Machinery, New York, NY, USA, 1232–1235. doi:10.1145/1056808.1056884

[8] Tara Boroushaki, Maisy Lam, Laura Dodds, Aline Eid, and Fadel Adib. 2023. Augmenting Augmented Reality with Non-Line-of-Sight Perception. In *20th USENIX Symposium on Networked Systems Design and Implementation (NSDI 23)*. USENIX Association, Boston, MA, 1341–1358. https://www.usenix.org/conference/nsdi23/presentation/boroushaki

[9] Paul Brandl, Tomislav Jukić, Reinhard Enne, Kerstin Schneider-Hornstein, and Horst Zimmermann. 2016. Optical Wireless APD Receiver With High Background-Light Immunity for Increased Communication Distances. *IEEE Journal of Solid-State Circuits* 51, 7 (2016), 1663–1673. doi:10.1109/JSSC.2016.2557815

[10] Christopher J. Breshike, Christopher A. Kendziora, Robert Furstenberg, Viet Nguyen, Andrew Kusterbeck, and R. Andrew McGill. 2019. Infrared backscatter imaging spectroscopy of trace analytes at standoff. *Journal of Applied Physics* 125, 10 (03 2019), 104901. doi:10.1063/1.5079622

[11] Charles J. Carver, Hadleigh Schwartz, Qijia Shao, Nicholas Shade, Joseph P. Lazzaro, Xiaoxin Wang, Jifeng Liu, Eric R. Fossum, and Xia Zhou. 2023. Catch Me If You Can: Demonstrating Laser Tethering with Highly Mobile Targets. In *Proceedings of the 29th Annual International Conference on Mobile Computing and Networking* (Madrid, Spain) *(ACM MobiCom '23)*. Association for Computing Machinery, New York, NY, USA, Article 108, 3 pages. doi:10.1145/3570361.3614081

[12] Kishan Chandan, Vidisha Kudalkar, Xiang Li, and Shiqi Zhang. 2021. ARROCH: Augmented Reality for Robots Collaborating with a Human. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Xi'an, China, 3787–3793. doi:10.1109/ICRA48506.2021.9561144

[13] Gaige Chen, Pei Wang, Bo Feng, Yihui Li, and Dekun Liu. 2020. The framework design of smart factory in discrete manufacturing industry based on cyber-physical system. *International Journal of Computer Integrated Manufacturing* 33, 1 (2020), 79–101. doi:10.1080/0951192X.2019.1699254

[14] D. Cook and S.K. Das. 2004. *Smart Environments: Technology, Protocols, and Applications*. John Wiley & Sons, Hoboken, New Jersey, USA.

[15] Denis C. Daly and Anantha P. Chandrakasan. 2007. An Energy-Efficient OOK Transceiver for Wireless Sensor Networks. *IEEE Journal of Solid-State Circuits* 42, 5 (2007), 1003–1011. doi:10.1109/JSSC.2007.894323

[16] Mustafa Doga Dogan, Raul Garcia-Martin, Patrick William Haertel, Jamison John O'Keefe, Ahmad Taka, Akarsh Aurora, Raul Sanchez-Reillo, and Stefanie Mueller. 2023. BrightMarker: 3D Printed Fluorescent Markers for Object Tracking. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 55, 13 pages. doi:10.1145/3586183.3606758

[17] Mustafa Doga Dogan, Eric J Gonzalez, Karan Ahuja, Ruofei Du, Andrea Colaço, Johnny Lee, Mar Gonzalez-Franco, and David Kim. 2024. Augmented Object Intelligence with XR-Objects. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 19, 15 pages. doi:10.1145/3654777.3676379

[18] Mustafa Doga Dogan, Ahmad Taka, Michael Lu, Yunyi Zhu, Akshat Kumar, Aakar Gupta, and Stefanie Mueller. 2022. InfraredTags: Embedding Invisible AR Markers and Barcodes Using Low-Cost, Infrared-Based 3D Printing and Imaging Tools. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 269, 12 pages. doi:10.1145/3491102.3501951

[19] Ruofei Du, Alex Olwal, Mathieu Le Goc, Shengzhi Wu, Danhang Tang, Yinda Zhang, Jun Zhang, David Joseph Tan, Federico Tombari, and David Kim. 2022. Opportunistic Interfaces for Augmented Reality: Transforming Everyday Objects into Tangible 6DoF Interfaces Using Ad hoc UI. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 183, 4 pages. doi:10.1145/3491101.3519911

[20] Anna Maria Feit, Shane Williams, Arturo Toledo, Ann Paradiso, Harish Kulkarni, Shaun Kane, and Meredith Ringel Morris. 2017. Toward Everyday Gaze Input: Accuracy and Precision of Eye Tracking and Implications for Design. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 1118–1130. doi:10.1145/3025453.3025599

[21] Danilo Gasques, Janet G. Johnson, Tommy Sharkey, and Nadir Weibel. 2019. What You Sketch Is What You Get: Quick and Easy Augmented Reality Prototyping with PintAR. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3290607.3312847

[22] Christopher Getschmann and Florian Echtler. 2021. Seedmarkers: Embeddable Markers for Physical Objects. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction* (Salzburg, Austria) *(TEI '21)*. Association for Computing Machinery, New York, NY, USA, Article 26, 11 pages. doi:10.1145/3430524.3440645

[23] Himanshu Gupta, Max Curran, Jon Longtin, Torin Rockwell, Kai Zheng, and Mallesham Dasari. 2022. Cyclops: an FSO-based wireless link for VR headsets. In *Proceedings of the ACM SIGCOMM 2022 Conference* (Amsterdam, Netherlands) *(SIGCOMM '22)*. Association for Computing Machinery, New York, NY, USA, 601–614. doi:10.1145/3544216.3544255

[24] Rakel Haakegaard and Joanna Lang. 2005. *The elliptic curve diffie-hellman (ecdh)*. University of California, Santa Barbara. Retrieved Feburary 2, 2025 from https://koclab.cs.ucsb.edu/teaching/ecc/project/2015Projects/Haakegaard+Lang.pdf

[25] Charles Herder, Meng-Day Yu, Farinaz Koushanfar, and Srinivas Devadas. 2014. Physical Unclonable Functions and Applications: A Tutorial. *Proc. IEEE* 102, 8 (2014), 1126–1141. doi:10.1109/JPROC.2014.2320516

[26] Sebastian Hubenschmid, Johannes Zagermann, Simon Butscher, and Harald Reiterer. 2021. STREAM: Exploring the Combination of Spatially-Aware Tablets with Augmented Reality Head-Mounted Displays for Immersive Analytics. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*

(Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 469, 14 pages. doi:10.1145/3411764.3445298

[27] Keiichi Ihara, Mehrad Faridan, Ayumi Ichikawa, Ikkaku Kawaguchi, and Ryo Suzuki. 2023. HoloBots: Augmenting Holographic Telepresence with Mobile Robots for Tangible Remote Collaboration in Mixed Reality. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology* (San Francisco, CA, USA) *(UIST '23)*. Association for Computing Machinery, New York, NY, USA, Article 119, 12 pages. doi:10.1145/3586183.3606727

[28] Allison Jing, Kieran William May, Mahnoor Naeem, Gun Lee, and Mark Billinghurst. 2021. eyemR-Vis: Using Bi-Directional Gaze Behavioural Cues to Improve Mixed Reality Remote Collaboration. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA '21)*. Association for Computing Machinery, New York, NY, USA, Article 283, 7 pages. doi:10.1145/3411763.3451844

[29] Hiroki Kaimoto, Kyzyl Monteiro, Mehrad Faridan, Jiatong Li, Samin Farajian, Yasuaki Kakehi, Ken Nakagaki, and Ryo Suzuki. 2022. Sketched Reality: Sketching Bi-Directional Interactions Between Virtual and Physical Worlds with AR and Actuated Tangible UI. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 1, 12 pages. doi:10.1145/3526113.3545626

[30] Jonathan Katz. 2010. *Digital Signatures*. Vol. 1. Springer, New York, NY, USA.

[31] Maruchi Kim, Antonio Glenn, Bandhav Veluri, Yunseo Lee, Eyoel Gebre, Aditya Bagaria, Shwetak Patel, and Shyamnath Gollakota. 2024. IRIS: Wireless ring for vision-based smart home interaction. In *Proceedings of the 37th Annual ACM Symposium on User Interface Software and Technology* (Pittsburgh, PA, USA) *(UIST '24)*. Association for Computing Machinery, New York, NY, USA, Article 114, 16 pages. doi:10.1145/3654777.3676327

[32] Juhee Lee. 2022. Problem-based gaming via an augmented reality mobile game and a printed game in foreign language education. *Education and Information Technologies* 27, 1 (2022), 743–771. doi:10.1007/s10639-020-10391-1

[33] Sueyoon Lee, Abdallah El Ali, Maarten Wijntjes, and Pablo Cesar. 2022. Understanding and Designing Avatar Biosignal Visualizations for Social Virtual Reality Entertainment. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 425, 15 pages. doi:10.1145/3491102.3517451

[34] Chengxi Li, Pai Zheng, Shufei Li, Yatming Pang, and Carman K.M. Lee. 2022. AR-assisted digital twin-enabled robot collaborative manufacturing system with human-in-the-loop. *Robotics and Computer-Integrated Manufacturing* 76 (2022), 102321. doi:10.1016/j.rcim.2022.102321

[35] Jieun Lim, Youngji Koh, Auk Kim, and Uichin Lee. 2024. Exploring Context-Aware Mental Health Self-Tracking Using Multimodal Smart Speakers in Home Environments. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 699, 18 pages. doi:10.1145/3613904.3642846

[36] Pin-Chun Lu, Che-Wei Wang, Yu Lun Hsu, Alvaro Lopez, Ching-Yi Tsai, Chiao-Ju Chang, Wei Tian Mireille Tan, Li-Chun Lu, and Mike Y. Chen. 2024. VeeR: Exploring the Feasibility of Deliberately Designing VR Motion that Diverges from Mundane, Everyday Physical Motion to Create More Entertaining VR Experiences. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 424, 13 pages. doi:10.1145/3613904.3642064

[37] Hongshen Ma and Joseph A Paradiso. 2002. The FindIT Flashlight: Responsive Tagging Based on Optically Triggered Microprocessor Wakeup. In *UbiComp 2002: Ubiquitous Computing*, Gaetano Borriello and Lars Erik Holmquist (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 160–167. doi:10.1007/3-540-45809-3_12

[38] João G. Martins, Gabriel M. Costa, Marcelo R. Petry, Paulo Costa, and António Paulo Moreira. 2024. A Multi-User Multi-Robot Collaboration through Augmented Reality. In *2024 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)*. IEEE, Paredes de Coura, Portugal, 144–150. doi:10.1109/ICARSC61747.2024.10535948

[39] Thomas D. Novlan, Harpreet S. Dhillon, and Jeffrey G. Andrews. 2013. Analytical Modeling of Uplink Cellular Networks. *IEEE Transactions on Wireless Communications* 12, 6 (2013), 2669–2679. doi:10.1109/TWC.2013.050613.120325

[40] Esben Warming Pedersen and Kasper Hornbæk. 2011. Tangible bots: interaction with active tangibles in tabletop interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. Association for Computing Machinery, New York, NY, USA, 2975–2984. doi:10.1145/1978942.1979384

[41] David Pointcheval and Jacques Stern. 2000. Security Arguments for Digital Signatures and Blind Signatures. *Journal of Cryptology* 13, 3 (2000), 361–396. doi:10.1007/s001450010003

[42] Wasiu O. Popoola, Enrique Poves, and Harald Haas. 2012. Spatial Pulse Position Modulation for Optical Communications. *Journal of Lightwave Technology* 30, 18 (2012), 2948–2954. doi:10.1109/JLT.2012.2208940

[43] Shwetha Rajaram and Michael Nebeling. 2022. Paper Trail: An Immersive Authoring System for Augmented Reality Instructional Experiences. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 382, 16 pages. doi:10.1145/3491102.3517486

[44] K.M. Rudrappa, H. D. Maheshappa, C. Puttamadappa, K. Somashekar, and K.S. Venkatesh Prasad. 2009. Implementing RC5 protocol for remote control applications. In *2009 International Conference on Control, Automation, Communication and Energy Conservation*. IEEE, Perundurai, India, 1–6.

[45] Houssem Saidi, Emmanuel Dubois, and Marcos Serrano. 2021. HoloBar: Rapid Command Execution for Head-Worn AR Exploiting Around the Field-of-View Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 745, 17 pages. doi:10.1145/3411764.3445255

[46] Ravi S. Sandhu. 1997. *Role-based Access Control*. George Mason University. Retrieved Feburary 2, 2025 from https://profsandhu.com/articles/advcom/a98rbac.pdf

[47] Maximilian Speicher and Michael Nebeling. 2018. GestureWiz: A Human-Powered Gesture Design Environment for User Interface Prototypes. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (Montreal QC, Canada) *(CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–11. doi:10.1145/3173574.3173681

[48] Jannis Strecker, Khakim Akhunov, Federico Carbone, Kimberly García, Kenan Bektaş, Andres Gomez, Simon Mayer, and Kasim Sinan Yildirim. 2023. MR Object Identification and Interaction: Fusing Object Situation Information from Heterogeneous Sources. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 3, Article 124 (Sept. 2023), 26 pages. doi:10.1145/3610879

[49] Yong-Chin Tan, Sandeep R. Chandukala, and Srinivas K. Reddy. 2022. Augmented Reality in Retail and Its Impact on Sales. *Journal of Marketing* 86, 1 (2022), 48–66. doi:10.1177/0022242921995449

[50] Bruce H. Thomas. 2012. A survey of visual, mixed, and augmented reality gaming. *Comput. Entertain.* 10, 1, Article 3 (Dec. 2012), 33 pages. doi:10.1145/2381876.2381879

[51] Zhao Tian, Kevin Wright, and Xia Zhou. 2016. The darkLight rises: visible light communication in the dark. In *Proceedings of the 22th Annual International Conference on Mobile Computing and Networking* (New York City, New York) *(MobiCom '16)*. Association for Computing Machinery, New York, NY, USA, 2–15. doi:10.1145/2973750.2973772

[52] Nguyen Van Huynh, Dinh Thai Hoang, Xiao Lu, Dusit Niyato, Ping Wang, and Dong In Kim. 2018. Ambient Backscatter Communications: A Contemporary Survey. *IEEE Communications Surveys & Tutorials* 20, 4 (2018), 2889–2922. doi:10.1109/COMST.2018.2841964

[53] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. 2008. Pose tracking from natural features on mobile phones. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*. IEEE, Cambridge, UK, 125–134. doi:10.1109/ISMAR.2008.4637338

[54] Daniel Wagner, Gerhard Reitmayr, Alessandro Mulloni, Tom Drummond, and Dieter Schmalstieg. 2010. Real-Time Detection and Tracking for Augmented Reality on Mobile Phones. *IEEE Transactions on Visualization and Computer Graphics* 16, 3 (2010), 355–368. doi:10.1109/TVCG.2009.99

[55] D. Wagner and D. Schmalstieg. 2003. ARToolKit on the PocketPC platform. In *2003 IEEE International Augmented Reality Toolkit Workshop*. IEEE, Tokyo, Japan, 14–15. doi:10.1109/ART.2003.1320418

[56] Daniel Wagner and Dieter Schmalstieg. 2007. ARToolKitPlus for Pose Tracking on Mobile Devices. In *12th Computer Vision Winter Workshop 07*. Technische Universität Graz, Graz, Austria, 139–146. Computer Vision Winter Workshop : CVWW 2007, CVWW 2007 ; Conference date: 06-02-2007 Through 08-02-2007.

[57] Chao Wang, Stephan Hasler, Manuel Mühlig, Frank Joublin, Antonello Ceravola, Joerg Deigmoeller, and Lydia Fischer. 2020. Designing Interaction for Multi-agent System in an Office Environment. In *Proceedings of the Eighth International Workshop of Chinese CHI* (Honolulu, HI, USA) *(Chinese CHI '20)*. Association for Computing Machinery, New York, NY, USA, 86–88. doi:10.1145/3403676.3404349

[58] Purui Wang, Lilei Feng, Guojun Chen, Chenren Xu, Yue Wu, Kenuo Xu, Guobin Shen, Kuntai Du, Gang Huang, and Xuanzhe Liu. 2020. 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* (London, United Kingdom) *(MobiCom '20)*. Association for Computing Machinery, New York, NY, USA, Article 6, 13 pages. doi:10.1145/3372224.3380883

[59] Xue Wang and Yang Zhang. 2024. TextureSight: Texture Detection for Routine Activity Awareness with Wearable Laser Speckle Imaging. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 184 (Jan. 2024), 27 pages. doi:10.1145/3631413

[60] Zeyu Wang, Cuong Nguyen, Paul Asente, and Julie Dorsey. 2021. DistanciAR: Authoring Site-Specific Augmented Reality Experiences for Remote Environments. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 411, 12 pages. doi:10.1145/3411764.3445552

[61] W. Weber. 1978. Differential Encoding for Multiple Amplitude and Phase Shift Keying Systems. *IEEE Transactions on Communications* 26, 3 (1978), 385–391. doi:10.1109/TCOM.1978.1094074

[62] Marko Wolf and Timo Gendrullis. 2012. Design, Implementation, and Evaluation of a Vehicular Hardware Security Module. In *Information Security and Cryptology - ICISC 2011*, Howon Kim (Ed.). Springer Berlin Heidelberg, Berlin, Heidelberg, 302–318. doi:10.1007/978-3-642-31912-9_20

[63] Yue Wu, Purui Wang, Kenuo Xu, Lilei Feng, and Chenren Xu. 2020. Turboboosting Visible Light Backscatter Communication. In *Proceedings of the ACM SIGCOMM 2020 Conference* (Virtual Event, USA) *(SIGCOMM '20)*. Association for Computing Machinery, New York, NY, USA, 186–197. doi:10.1145/3387514.3406229

[64] Chang Xiao, Ryan Rossi, and Eunyee Koh. 2022. iMarker: Instant and True-to-scale AR with Invisible Markers. In *Adjunct Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22 Adjunct)*. Association for Computing Machinery, New York, NY, USA, Article 32, 3 pages. doi:10.1145/3526114.3558721

[65] Xieyang Xu, Yang Shen, Junrui Yang, Chenren Xu, Guobin Shen, Guojun Chen, and Yunzhe Ni. 2017. PassiveVLC: Enabling Practical Visible Light Backscatter Communication for Battery-free IoT Applications. In *Proceedings of the 23rd Annual International Conference on Mobile Computing and Networking* (Snowbird, Utah, USA) *(MobiCom '17)*. Association for Computing Machinery, New York, NY, USA, 180–192. doi:10.1145/3117811.3117843

[66] Yihui Yan, Zezhe Huang, Feiyang Xudu, and Zhice Yang. 2022. Enabling Tangible Interaction on Non-touch Displays with Optical Mouse Sensor and Visible Light Communication. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 109, 14 pages. doi:10.1145/3491102.3517666

[67] Zihan Yan, Yuxiaotong Lin, Guanyun Wang, Yu Cai, Peng Cao, Haipeng Mi, and Yang Zhang. 2023. LaserShoes: Low-Cost Ground Surface Detection Using Laser Speckle Imaging. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 853, 20 pages. doi:10.1145/3544548.3581344

[68] Jackie (Junrui) Yang and James A. Landay. 2019. InfoLED: Augmenting LED Indicator Lights for Device Positioning and Communication. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology* (New Orleans, LA, USA) *(UIST '19)*. Association for Computing Machinery, New York, NY, USA, 175–187. doi:10.1145/3332165.3347954

[69] Xiaoying Yang, Jacob Sayono, Jess Xu, Jiahao Nick Li, Josiah Hester, and Yang Zhang. 2022. MiniKers: Interaction-Powered Smart Environment Automation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 3, Article 149 (Sept. 2022), 22 pages. doi:10.1145/3550287

[70] Hui Ye and Hongbo Fu. 2022. ProGesAR: Mobile AR Prototyping for Proxemic and Gestural Interactions with Real-world IoT Enhanced Spaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 130, 14 pages. doi:10.1145/3491102.3517689

[71] Hui Ye, Jiaye Leng, Pengfei Xu, Karan Singh, and Hongbo Fu. 2024. ProInterAR: A Visual Programming Platform for Creating Immersive AR Interactions. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 610, 15 pages. doi:10.1145/3613904.3642527

[72] Steve Chi-Yin Yuen, Gallayanee Yaoyuneyong, and Erik Johnson. 2011. Augmented reality: An overview and five directions for AR in education. *Journal of Educational Technology Development and Exchange (JETDE)* 4, 1 (2011), 11. doi:10.18785/jetde.0401.10

[73] Tengxiang Zhang, Zitong Lan, Chenren Xu, Yanrong Li, and Yiqiang Chen. 2023. BLEselect: Gestural IoT Device Selection via Bluetooth Angle of Arrival Estimation from Smart Glasses. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 198 (Jan. 2023), 28 pages. doi:10.1145/3569482

[74] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 2022. StoryBuddy: A Human-AI Collaborative Chatbot for Parent-Child Interactive Storytelling with Flexible Parental Involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 218, 21 pages. doi:10.1145/3491102.3517479

[75] Fengyuan Zhu and Tovi Grossman. 2020. BISHARE: Exploring Bidirectional Interactions Between Smartphones and Head-Mounted Augmented Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3313831.3376233

[76] Zhengzhe Zhu, Ziyi Liu, Tianyi Wang, Youyou Zhang, Xun Qian, Pashin Farsak Raja, Ana Villanueva, and Karthik Ramani. 2022. MechARspace: An Authoring System Enabling Bidirectional Binding of Augmented Reality with Toys in Real-time. In *Proceedings of the 35th Annual ACM Symposium on User Interface Software and Technology* (Bend, OR, USA) *(UIST '22)*. Association for Computing Machinery, New York, NY, USA, Article 50, 16 pages. doi:10.1145/3526113.3545668