Camera-Free Object Discovery and Interaction with *GazeTags*

Vimal Mollyn
Carnegie Mellon University
Pittsburgh, PA, USA
vmollyn@cs.cmu.edu

Chris Harrison
Carnegie Mellon University
Pittsburgh, PA, USA
chris.harrison@cs.cmu.edu



Figure 1: GazeTags enable head-oriented gaze-aware environments and interactions *without cameras*. A narrow infrared signal emitted from a user's smart glasses (or earbuds) activates GazeTags the user is facing. In this example, a GazeTag attached to a potted plant reports back its object type (Ponytail Palm) and metadata (14% soil moisture). Using this information, a passively-monitoring AI agent could *proactively* prompt the user to water the plant. Or more conventionally, the user could also ask their AI agent if their plant needs water. At roughly $2 per GazeTag, which runs for more than a year on a coin cell battery, environments could be densely instrumented with smart functionality.

Abstract

We present GazeTags, a camera- and computer-vision-free approach for worn head-gaze object awareness. Our small electronic tags cost roughly $2 each and can operate for more than a year on a coin cell battery. They work in concert with a user-worn device (smart-glasses, hearables, etc.) that periodically emits a focused infrared signal aligned with the user's head direction. Only GazeTags within that narrow head-oriented beam wake and respond, providing their identity and optional payloads such as object state or live sensor readings. By avoiding cameras and computer vision pipelines, our approach is very energy efficient, able to operate continuously even in battery-constrained wearables, and sidesteps non-trivial privacy issues. We describe our hardware design, characterize our system performance and operating tradeoffs, and demonstrate how GazeTags can enable practical gaze-aware environments.

This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.
UIST '26, Detroit, MI, USA
© 2026 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-2856-3/2026/11
https://doi.org/10.1145/3830398.3830713

CCS Concepts

• **Human-centered computing** → **Ubiquitous and mobile computing**.

Keywords

Context awareness, smart environments, gaze, attention, IoT, visible-light communication, tags, AI agents, proactive assistance.

ACM Reference Format:
Vimal Mollyn and Chris Harrison. 2026. Camera-Free Object Discovery and Interaction with *GazeTags*. In *The 39th Annual ACM Symposium on User Interface Software and Technology (UIST '26), November 02–05, 2026, Detroit, MI, USA*. ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3830398.3830713

1 Introduction

Smart glasses, AI pins, earbuds and other head-worn devices offer the promise of always-available and proactive assistance. To achieve their full potential, these systems must understand what in the environment is relevant to the user. Gaze is one of the strongest signals for this purpose, often revealing a user's intent or task. This can be fine-grained eye-vector gaze (what specific object are they looking at) or coarser-grained head-vector gaze (what part of a scene are they facing towards). Systems with this information can adapt their user interfaces and offer timely assistance, such as



Figure 2: Example GazeTag uses. (A) A kettle equipped with a GazeTag and thermistor reports its current temperature. (B) A fire extinguisher provides its last serviced date to a health and safety inspector. (C) A user instructs a GazeTag equipped lamp to "turn off".

retrieving information about appliances, surfacing instructions for tools, or reacting to changes in attended objects.

Today, however, gaze-aware wearables generally rely on cameras and computer vision. While powerful, this is especially problematic for small, battery-constrained wearable platforms. Cameras are relatively power-hungry sensors, and vision pipelines are also energy expensive, either when run locally or requiring transmission elsewhere for processing. As a result, continuous operation remains an outstanding challenge for compact devices such as glasses, earbuds, and pins. In practice, this most often means that vision-based sensing is either explicitly triggered by wearers for short periods, or launched periodically with a low duty cycle, limiting its usefulness for always-available contextual assistance.

A second major challenge of vision-powered gaze for wearables is privacy. Even when such systems are technically effective, their social acceptability can be limited by discomfort around image capture, uncertainty about what is being recorded, and concern over downstream analysis. This tension is particularly acute for wearable AI, where the ideal technical solution is *persistent* vision sensing. This could include video capture in sensitive contexts like bathrooms, where AI assistance could be invaluable for some users. However, active cameras in these contexts are presently non-starters for many users and bystanders.

In this paper, we explore a different path: achieving head-oriented gaze sensing *without* cameras or computer vision, sidestepping many significant power and privacy issues. Our proposed solution is GazeTags, a low-cost tagging approach for instrumenting environments so that head-worn devices can discover and interact with attended objects. A user-worn device periodically emits a narrow field-of-view infrared signal that is naturally aligned with the user's head direction. Only tags within that head-gaze-correlated infrared "spotlight" wake and respond. In reply, tags announce both identity and dynamic payloads, such as object state, sensor readings, and contextual metadata. In this way, GazeTags can even move beyond computer-vision sensing, exposing non-visual states of objects, such as temperature, moisture, and power draw.

Of course, in order to be practically deployed across e.g., scores of objects in a home, GazeTags must be *ultra-low-cost*, *small*, and *maintenance-free* over long periods. Our implementation achieves these three properties. GazeTags cost around $2 USD each, including batteries, and could fall below $1 in high volumes. This makes dense instrumentation of everyday environments feasible (e.g., a home kitchen with 15 appliances would cost around $30 to augment). Second, GazeTags are small pucks, measuring 30 mm in diameter and 18 mm thick. They can be discreetly attached to objects of interest with magnets or tape. Third, GazeTags run for more than one year on a coin cell battery; when depleted, the battery can be replaced in seconds.

At a high level, our approach offers several advantages. Foremost, GazeTags do not require cameras, image processing, or computer vision models, and they are far better matched to the energy and compute budgets of small wearable devices. They also avoid many of the privacy concerns associated with always-on visual sensing, since no imagery needs to be captured or analyzed. Yet, GazeTags preserve a form of gaze-aware object discovery and interaction that is highly useful for wearable interfaces.

In the next section we highlight potential GazeTag capabilities through a set of example scenarios. We then review related work, focusing on other tag techniques and gaze sensing approaches. We detail our implementation, followed by a multi-part evaluation. Overall, results show our system works up to ranges of 8 m and with many GazeTags operating simultaneously.

2 Example GazeTag Uses

Although simple in design, GazeTags enable a variety of powerful interactions. To help illustrate this breadth of capability, we created a series of example uses (Figures 1 and 2). These are fully functional demos; please also see the Video Figure.

2.1 Object-Aware Assistants

We know from prior work that patterns of activity are powerful inputs to context-aware interfaces and AI assistants [67, 68]. Thus,

a simple yet powerful use case for GazeTags is to provide a real-time and continuous stream of what objects the user is facing. This could help an AI assistant stay informed about a user's current indoor location, task, and potentially intent: The user just entered the kitchen, just looked towards their coffee machine, etc.

2.2 Objects with Sensors & State

Beyond transmitting object identity, GazeTags can also interface with sensors. For example, a GazeTag attached to a plant pot could report its soil moisture level (Figure 1), allowing an AI assistant to proactively alert the user when it has been too long between waterings. Similarly, a GazeTag on a kettle could report its current temperature, allowing an AI assistant to gauge time to completion or alert the user when the water has reached an ideal temperature for tea vs. coffee (Figure 2A).

2.3 Personalization & Access Control

GazeTags can also tailor their response based on who is looking at them (user-worn devices transmit user credentials). One option is for GazeTags to only respond to their owner. Alternatively, GazeTags could reply with different information to different sets of users. As one example, a GazeTag on a fire extinguisher could report its object type to anyone, but include its last service date when seen by health and safety personnel (Figure 2B).

2.4 Object-Integrated GazeTags with Actuation

If integrated into objects, GazeTags could allow for actuation. For example, after reading all GazeTags in a scene, a user-device could follow up with a command, either triggered explicitly by a user or by an AI agent on behalf of the user. As a simple example, a user could look at a lamp and say "turn off." The user-worn device would transmit the recipient object's UID and command, and the GazeTag attached to the lamp would toggle it off (Figure 2C).

3 Related Work

In this section, we review related work on sensing user context. We organize by approach, first reviewing work that primarily uses computer vision, then systems that instrument the environment with passive tags, and conclude with active tagging approaches, which are most similar to GazeTags.

3.1 Sensing Context from Vision

Cameras are perhaps the most common sensing modality for understanding user context, being widely used in both research [43, 51, 62] and commercial systems [28, 49]. They can infer context from a user's gaze direction [16, 29, 34, 36, 38, 56], their body and hand pose [27, 53, 54] and any objects attended [14, 15, 57]. Pointing [11, 19, 33, 39] and head-gaze [44] can further inform where in the world a user is referring to. More recently, cameras have also been used to infer user intent [7, 20, 37, 43, 62], e.g., where are the user's hands moving next and what are they going to grab? While powerful, these approaches are power hungry, computationally expensive and privacy invasive, limiting their use on a battery-constrained, always-on wearable.

3.2 Sensing Context with Passive Tags

Another approach for sensing context is to instrument the environment with tags, read by a user-worn device. Tags can offer the benefit of increased identification accuracy, data-bearing payloads, and the ability to resolve visually identical objects (by using different IDs). We first discuss passive tags, which are either unpowered or receiving power from a user-worn transmitter.

The most pervasive passive tagging technology is barcodes, which can be read by a camera or close-range optical scanner. Later standards, such as AprilTags [48], ArUco [25], and ARTags [21] can reveal not only identity, but also 6DoF pose, allowing for spatial interactions. However, similar to other vision-based approaches, they require a camera and computer vision pipeline, which can be power hungry and raise privacy concerns. Moreover, to work at longer ranges, tags have to be unreasonably large (e.g., >10x10cm) so as to be visually obtrusive, or require very high resolution cameras.

Another popular approach is to use passive RFID tags, which can communicate through RF backscatter [13, 58, 64] and have been used to detect activity context [8, 12, 40, 41, 63]. Unfortunately, RFID readers (which transmit) are generally large, activate a wide area, are limited in range to a few meters, and are energy expensive. For these reasons, they have not been integrated into wearables, and the range precludes them from the types of use we envision for GazeTags. NFC technology [17, 64, 75] is highly related, and small, low-powered readers have been integrated into wearables. However, as the name suggests, NFC tags are near field, operating at centimeter or body scales, and thus would not work for the types of interactions we describe.

3.3 Sensing Context with Active Tags

It is also possible to instrument the environment with active tags for context sensing. Popular wireless tag protocols include Bluetooth [5, 42], ZigBee [9, 26], and Wi-Fi [6, 76]. Approximate distance to these tags can be estimated via RSSI and similar signals. Other approaches have been explored in the literature, including powered tags that communicate wirelessly through modulated light [3, 50, 52, 69, 72], illuminated fiducial markers [47], and ultrasound [55, 70]. In general, these are very energetic approaches that are not compatible with long term deployment on small batteries. To avoid this significant battery issue, researchers have also explored tags that generate their own energy through object actuation and interaction [24, 73, 74, 78].

Among all prior work, the most closely related to GazeTags in terms of capability are UWB (ultra-wideband) tags, which exist as commercial products (e.g., Apple AirTags, Samsung SmartTags). These tags could, in principle, support head-gaze-aware interaction at room-scale ranges and can be made relatively low power [4, 18, 79]. The main downside is UWB remains expensive even after many years of commercial optimization. The bill of materials for an Apple AirTag is estimated at roughly $10 (with a retail cost of $29) [31]; the AirTag's Nordic SoC alone costs approximately $4, even at large volumes [59]. At 5× the cost of our GazeTags, dense deployments of hundreds of UWB tags across e.g., a household is not yet consumer feasible. Secondarily, although UWB needs only compact simple antennas for range-finding (as it relies on time of flight, not antenna directivity), multi-element array antennas are needed to estimate

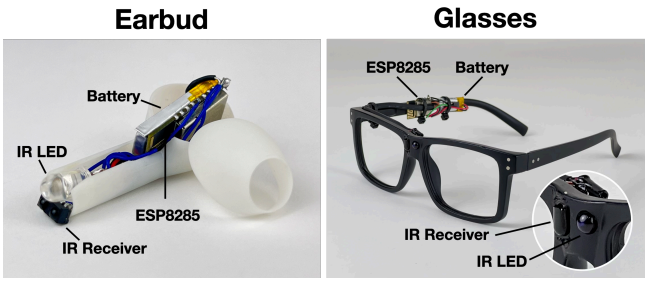


Figure 3: We created two head-worn prototypes to demonstrate the feasibility of GazeTags.

Part	Subtotal
Microcontroller (ATtiny85)	$0.41
Infrared Receiver (TSOP18338)	$0.41
MOSFET (2N7000, QTY: 2)	$0.08
Misc. Passives	$0.04
Infrared LED (140° FoV)	$0.29
Battery (CR2477, 1100 mAh)	$0.60
Battery Holder	$0.26
PCB	$0.06
Enclosure	$0.05
Total	**$2.20**

Table 1: Bill of materials for a GazeTag. When applicable, subtotals are based on purchase volumes of 10K units.

angle-of-arrival (AoA) on the reader device. Such antennas have yet to be demonstrated in small worn devices, such as earbuds or glasses (at present, the smallest UWB AoA implementation is in smartphones). Finally, UWB tags, like other RF approaches, can communicate without line of sight, and tags can be behind walls, inside cabinets, etc. This behavior is desirable when trying to find e.g., the key chain you lost in your house. However, it is undesirable behavior for our use case, where we wish to only know objects that lie in a user's field of view to contextually assist them.

3.4 Infrared Communication

GazeTags utilize infrared (IR) communication, which has a long history as a simple, low-cost, power-efficient, wireless channel for short-range digital communication. In consumer electronics, IR has been used for many decades in remote controls for televisions, stereos, air conditioners, and related appliances. Beyond one-way command input, infrared also supported richer bidirectional device-to-device communication through the Infrared Data Association (IrDA) standards [30, 46], which emerged in the mid-1990s and were widely adopted in laptops, PDAs, phones, printers, and cameras for short-range data exchange.

Rather than introducing a new wireless substrate, GazeTags leverage existing infrared communication components. This brings several practical advantages: IR components are extremely low cost, highly available, and robust under everyday operating conditions, including bright light. As a result, GazeTags can inherit the reliability and affordability of an established technological ecosystem, repurposing it for gaze-aware object discovery.

We note that IR has previously been explored for identifying and selecting physical devices. Early among these is *That one there!* [66], which used an IR-emitting stylus to point at devices, enabling two-way identification and subsequent wireless connection establishment. More closely related, HOBS [77] used a glasses-worn IR transmitter to select IR-tagged objects based on head orientation. Like GazeTags, HOBS exploits the directionality of IR to associate a wearer's head orientation with physical objects in the environment, and further explored techniques for refining selection when multiple tags fall within the IR field of view. Also very similar, InSight [10] used a glasses-worn IR transmitter (G-Ripple) and object-attached IR receiver tags (G-Sens) to pair arbitrary input controllers to smart devices such that the user could control objects they were facing. Each G-Sens "tag" device lasted only 50 hours on a single charge. We do not claim novelty in using head-directed IR for tagged-object selection. Instead, GazeTags explores a different point in this systems design space: whether such an approach can support dense, persistent instrumentation using tags that are ultra-low-cost, long-lived, small, and capable of reliably returning both identity and dynamic object state. Moreover, rather than focusing primarily on explicit object selection, we envision GazeTags continuously and passively exposing contextual object information to worn AI agents, which can use these signals to infer context and proactively support users' ongoing tasks.

4 Implementation

We now describe the hardware and software implementation of GazeTags and the user-worn devices that communicate with them.

4.1 Head-Worn Device Prototype

To demonstrate that GazeTags can work with a variety of head-worn form factors, we built two prototype devices: smart glasses and earbuds (Figure 3). Both use the same core hardware — an ESP8285 SoC [2], a TSOP18338 infrared receiver [61], and a 940 nm infrared LED. The latter wavelength was purposely selected as it is completely outside the visible spectrum (unlike common 790 or 850 nm infrared LEDs). Also, like previous infrared communication systems, we used modulated carriers (38 kHz) which offer robust operation under ambient light. We discuss selection of LED beam angle in the next section. The ESP8285 SoC was selected for its small footprint ($20.0 \times 12.3 \times 3.0$ mm) and built-in wireless connectivity.

Unlike our GazeTags, which must operate for long durations without maintenance to be practical, worn devices are typically charged daily. This significantly relaxes their power constraints, and thus our head-worn devices take on the more energy-expensive role of transmitting a wake-up signal. Specifically, to wake GazeTags, a head-worn device emits a 300 ms wake-up signal consisting of a 300 μs 38 kHz square wave (50% duty cycle) followed by a 700 μs blanking interval (Figure 5A). It then transmits a GazeTag signal with optional user credentials and waits for a reply from nearby tags. This transmission consumes 0.007 mWh of a head-worn device's battery (this is total energy usage, including the LED, ESP8285, and IR receiver). For reference, Meta Ray-Ban Display glasses contain a 960 mWh battery [45]. Of course, this wake-up signal would only trigger periodically, e.g., every 3 seconds, or be intermittently

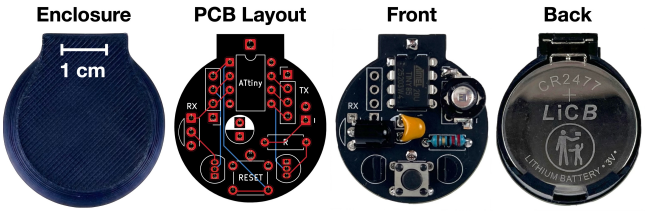


Figure 4: Different views of a GazeTag.

triggered by an AI assistant when it determines a user might benefit from contextual assistance.

4.2 Head-Worn Narrow Beam Transmitter

A tradeoff we make being camera-free is that we cannot track eye gaze, and must fall back to a less accurate gaze proxy: head direction. Fortunately, prior work has shown that users' eyes most often stay within a comfortable range of ±25° [65], and a majority of the time, eyes deviate from the head vector less than ±20° [23]. This informed the selection of our head-worn infrared transmitter: We selected an LED with a ±20° emission pattern. As LEDs are almost point light sources, the beam shape is roughly that of a cone.

4.3 GazeTag Hardware

We designed our GazeTags (Figure 4) around the CR2477 coin cell battery, which measures 24.5 mm in diameter and 7.7 mm in height. The CR2477 provides 1100 mAh of capacity, which provides more than a year of runtime (discussed in Section 4.5). We fabricated a custom PCB sized to match the CR2477's footprint (Figure 4). The board hosts an ATtiny85 microcontroller, which we selected for its excellent deep sleep modes and low cost. The ATtiny85 drives a 940 nm infrared LED using a 2N7000 MOSFET. Unlike the narrower, cone-shaped beam LED used in the headset (Section 4.2), the GazeTag uses a 140° conical emission LED. This wide emission pattern provides a large effective viewing range, regardless of how the tag is oriented on a surface or object. The tag uses the same TSOP18338 infrared receiver as the headset [61], but switches it on and off to conserve power, using another 2N7000 MOSFET (discussed in Section 4.5). The front is enclosed in a friction-fit, 3D-printed case using an infrared-transparent filament [1], allowing the LED and receiver to operate without cutouts. In total, each GazeTag costs $2.20; a bill of materials is provided in Table 1.

4.4 Communication & GazeTag Firmware

In this section, we provide an overview of how GazeTags and the aforementioned user-worn device communicate. As we step through this process, we also explain relevant GazeTag implementation details. Power consumption is discussed in Section 4.5.

4.4.1 Wake-up Interrupt. GazeTags wake up from deep sleep (Figure 5I) every 130 ms for 7 ms (i.e., 5.1% duty cycle). During this brief interval, the GazeTag powers on its IR receiver. For the first 1 millisecond, the IR receiver is allowed to stabilize its internal automatic gain control (AGC). We then enable microcontroller interrupts on the IR receiver's digital pin and wait 6 ms. If no IR transmission is detected during this period, interrupts are disabled, the IR receiver is powered down, and the microcontroller goes back to deep sleep (Figure 5I). If an IR transmission is detected, triggering an interrupt (Figure 5B), we wait 350 ms to ensure the head-worn device has completed its wake-up signal (Figure 5A) and finished transmitting any data (Figure 5C/E). During this interval, interrupt handlers are used to buffer incoming IR data. Each time data is appended to the buffer, it is checked to see if it contains a valid transmission. If one is detected and correctly decoded, we set a flag for later processing. For all of our communication, we use the FAST IR protocol [32].

4.4.2 Head-Worn Device Sender Check. In the IR receiver interrupt handler, GazeTags check for a valid sending device signal (Figure 5D). This is a single FAST packet containing a fixed 8 bit ID — we selected 0x68 ('h' for head device). This is necessary so that GazeTags do not reply to other IR transmissions in the environment, such as TV remote controls. Not only would this waste valuable battery power, but it would also interfere with the operation of other devices. If the IR transmission was not from an applicable GazeTag-capable device, the tag simply returns to deep sleep (Figure 5I).

4.4.3 Access Control (Optional). The head-worn device can also transmit a user ID (Figure 5E), allowing only authorized users to trigger a GazeTag response. Our current prototype uses a hard-coded UID as a simple access-control mechanism (Figure 5F). This is sufficient for proof-of-concept use, but it is not secure in practice: a nearby adversary could observe the UID and replay it to impersonate the owner. A full implementation would use a rolling credential or lightweight cryptographic challenge-response protocol, preventing naive sniff-and-replay attacks.

4.4.4 Reply Payload. The default reply of a GazeTag is a 16 bit object UID contained in two back-to-back packets, each taking 29 ms to transmit [32]. A user's head worn device could maintain their own UID-to-device lookup tables, or this could be a cloud service. A range of UIDs could be reserved to identify common objects (e.g., 0x4C9E = any water fountains), while a separate range could be reserved for personalized one-to-one mappings (e.g., 0xA7F3 = water fountain, Newell Simon Hall, 1st floor).

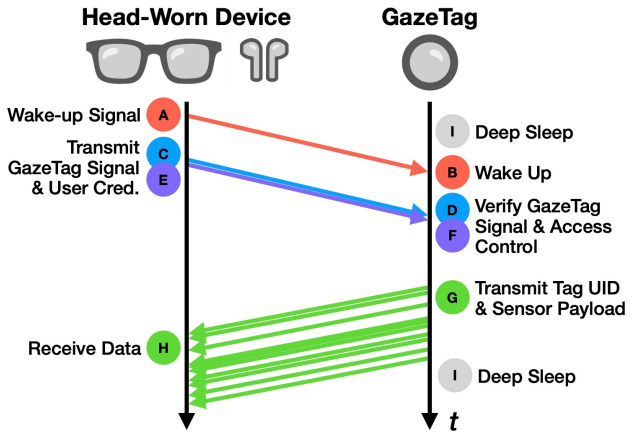


Figure 5: Overview of the communication sequence between a head-worn device and a GazeTag. Time not to scale.

4.4.5 Sensor Data (Optional). The microcontroller we use for our GazeTags has two unused pins that we expose as through-hole pads on our PCB. Both pins are capable of reading analog values, or interfacing with digital sensors. We describe some example uses for sensor data in Section 2. When we want to reply with sensor data, we append an extra FAST packet to our reply payload.

4.4.6 Multi-Tag Collision Avoidance. If more than one GazeTag is in the receptive field of the head-worn device, multiple tags will compete to reply in a naive implementation, potentially leading to a subset or even no tags being recognized. To avoid collisions and ensure high detection reliability, we employ two schemes. First, GazeTags emit ten replies with random delays between 40 and 200 ms (Figure 5G). Each 16-bit reply takes 58 ms to transmit, and so collisions are comparatively rare when there are just a few tags. Second, GazeTags implement carrier-sense multiple access (CSMA). Put simply, after waking from a random delay, GazeTags check to see if any other IR transmissions are currently active (i.e., carrier sense), and if so, go back to sleep anywhere from 30 to 50 ms to try transmitting later. This simple check significantly decreases collisions and increases reliability.

To help inform the main variables of our collision avoidance implementation, namely number of reply repeats and random delay range, we created a Monte Carlo simulation tool. Plots of simulated data (with and without CSMA) are included in Figure 6. To achieve a ≥99% success rate with 5 co-located tags, we selected 10 repeats, with a 200 ms max delay, with CSMA. We evaluate real-world performance with up to 6 GazeTags side-by-side in our later study (Section 5.3).

4.5 GazeTag Power Consumption & Battery Life

We drew design inspiration from Apple AirTags, which last roughly one year with typical use and have a user-swappable coin cell battery. Similarly, we targeted one-year run-times with a replaceable coin cell battery. To achieve such long battery life, we had to significantly optimize our device's power consumption. For this, we made extensive use of a Nordic Power Profiler Kit II [60].

One easy optimization was down-clocking the ATtiny to 1 MHz, which still offers plenty of compute for our needs. Our device spends ~94.9% of its time in deep sleep, during which it consumes just 4.7 μA. The IR receiver (TSOP18338) consumes 0.67 mA when continuously powered, and for this reason, we only wake to power it on intermittently to check for incoming IR transmissions. This accounts for 5.1% of device time, during which the tag consumes 1.6 mA. Amortized, this means we have a mean continuous power draw of 86.1 μA. We use a CR2477 coin cell battery with a rated capacity of 1100 mAh and a nominal voltage of 3.0 V. This means our GazeTags have a maximum operational life of ~1.4 years.

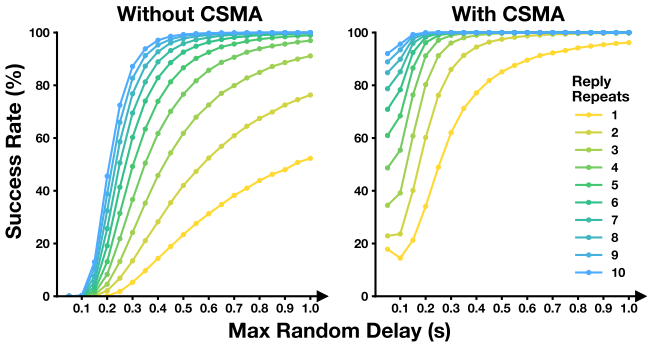


Figure 6: To inform our GazeTag collision avoidance implementation, we ran a Monte Carlo simulation with five co-located GazeTags. We vary the number of replies, delays between replies, and with or without CSMA.

Of course, power is also consumed when the tag has to reply by turning on its IR LEDs, which is a comparatively rare, but power-expensive operation. The LED, when consistently on, consumes 74.7 mA. FAST packets are comprised of a 3.2 ms header, 25.2 ms payload, and 0.5 ms terminating mark [32] (29 ms total). IR transmissions are 50% duty cycle, 38 kHz PWM signals, and on average 50% of the payload bits are zeros (LED off, consuming no power), so the effective power draw is closer to 22.3 mA during transmission.

As already described above, to ensure reliable transmission even in the presence of other tags and IR systems, GazeTags transmit ten replies (580 ms total transmit time). Thus a single full reply consumes 0.0062 mAh (all components combined). If we assume an average usage of twenty interactions every day for a year, this would consume 4.1% of the total capacity of the battery (1100 mAh).

5 Evaluation

To evaluate the performance of GazeTags, we conducted three experiments examining the head-worn device's directionality, GazeTags directionality, and multi-tag detection. Figure 7 provides an illustration of the three experimental setups. In all three experiments, we evaluate end-to-end accuracy (i.e., round-trip communication performance) across three challenging long ranges: 4, 6, and 8 m.

5.1 Head-Worn Device Directionality & Range

As already discussed, for GazeTags to serve as a reliable proxy for head gaze, the head-worn infrared transmitter must have a sufficiently narrow and directional beam. To systematically characterize the transmitter's directionality, we conducted an automated experiment. Specifically, we mounted our head-worn transmitter on an SO-101 robot arm [35] at a height of 1 m (Figure 7). A single GazeTag was placed on a tripod at the same height, directly facing the transmitter. We tested three distances between the transmitter and tag: 4 m, 6 m, and 8 m, chosen to represent typical interaction distances in indoor environments such as homes and offices. We intentionally excluded shorter distances (e.g., 1–2 m), as detection at close range is expected to be near-perfect, and we wanted to test the system at more challenging ranges.

We controlled the robot arm and head-worn device with a Python script. The robot arm swept through yaw and pitch angles ranging from −80° to +80° in 10° increments. At each orientation, the head-worn device emitted a wake-up signal followed by a GazeTag signal and waited for a reply (3500 ms timeout). As described in Section 4.4, each GazeTag responds with up to ten reply transmissions to help overcome collisions with other infrared transmitting devices, including potentially other GazeTags in the environment. A trial was counted as successful if at least one reply was received

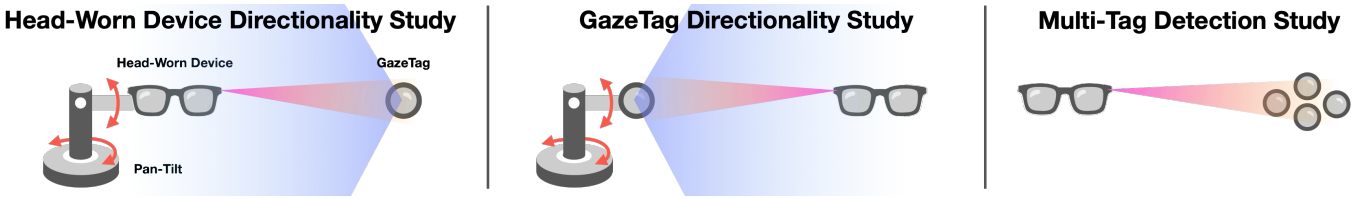


Figure 7: Illustration of the experimental setup for our three performance studies. Not to scale. Each study was repeated at 4, 6 and 8 m distances to also evaluate range. Results plotted in Figures 8 and 9.

by the head-worn device within the window. This process was repeated 10 times at each pitch–yaw–distance combination, yielding $17 \times 17 \times 10 \times 3 = 8{,}670$ total trials. The study was conducted indoors in typical office lighting and took approximately 4 hours to complete per distance.

5.2 GazeTag Directionality & Range

Users may approach a GazeTag from a wide range of angles, and thus, for reliable operation, GazeTags must operate with a wide field of view. This not only includes receiving the head-worn device's wake-up signal, but also for transmitting a reply back to the sender (enabled by the wide-angle LED described in Section 4.3). To systematically characterize our tag's receptive field, we conducted a second automated experiment.

We used the same apparatus and procedure as described in Section 5.1, but with the positions of the GazeTag and head-worn device swapped. Specifically, we mounted a GazeTag on the robot arm and placed the head-worn device on a tripod at the same height (1 m), directly facing the tag. As before, a Python script controlled the robot arm and head-worn device. The robot arm rotated the GazeTag through yaw and pitch angles from $-80°$ to $+80°$ in $10°$ increments at distances of 4 m, 6 m, and 8 m. The communication protocol, success criterion, and number of repetitions were identical, again yielding $17 \times 17 \times 10 \times 3 = 8{,}670$ total trials. This study also took approximately 4 hours to complete per distance.

5.3 Multi-Tag Detection Study

In real-world use, multiple GazeTags may be within the field of view of the head-worn device, especially at longer distances. To test the efficacy of our collision avoidance implementation (Section 4.4.6), we conducted a third experiment. For this, we placed the head-worn device on a tripod pointed directly at a 25 cm wide shelf at the same height (1 m). We evenly spaced between 2 and 6 GazeTags on the shelf, with each tag programmed to transmit its own unique UID. As already discussed in Section 4.4.6, we optimized our design to support 5 simultaneous GazeTags. Thus, we selected one above this quantity to see how accuracy would degrade. As with our previous studies, we tested at distances of 4 m, 6 m, and 8 m. The head-worn device emitted a wake-up signal followed by a GazeTag signal, and waited up to 3.5 seconds for any replies. We counted a trial as successful if at least one reply was received from every tag present. Each combination of tag count and distance was repeated 100 times. The study took a total of roughly 3 hours to complete.

6 Results

A summary of results can be found in Figures 8–9. Importantly, all accuracies reported in this section are computed as successful round-trip communications (i.e., head-worn device transmitting, GazeTag receiving, GazeTag transmitting, head-worn device receiving). At the high level, we found our system works well up to a range of 6 m, and operates in excess of 99% reliably even with five GazeTags present within a small, 25 cm area.

6.1 Head-Worn Device Directionality & Range

Results from our head-worn device directionality study are visualized in Figure 8, which characterizes the beam shape. Within the

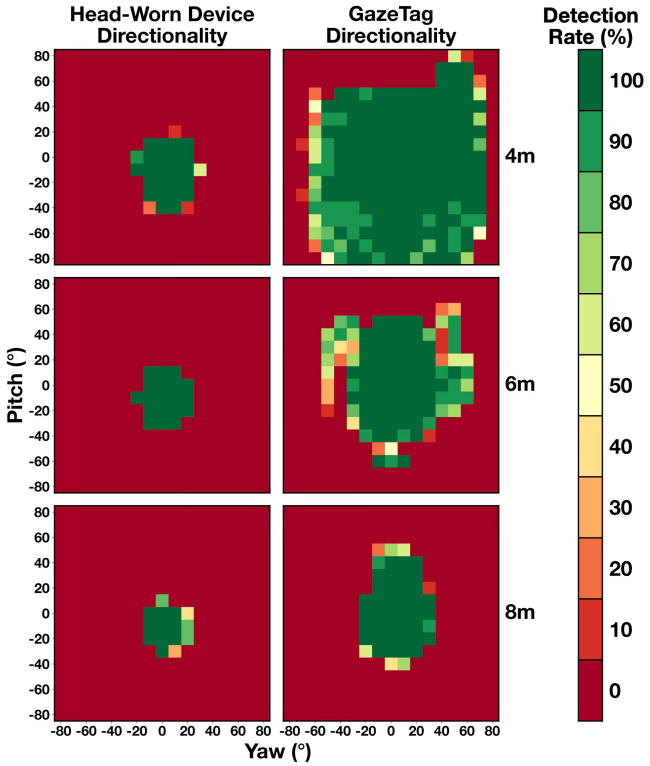


Figure 8: Results of our two directionality studies (see also Figure 7, left and center illustrations). Note the head-worn device has a purposely narrow beam shape, while GazeTags have a purposely wider one.

intended narrow beam, detection accuracy is near 100%. There is a slight narrowing of the effective field of view as the range grows. This is because although the beam spreads linearly with distance, the light intensity falls off as the inverse square of distance. However, across all ranges tested, our head-worn device maintains a usefully narrow activation "spotlight" onto the scene.

6.2 GazeTag Directionality & Range

Results from our tag directionality study are visualized in Figure 8. At 4 m, the GazeTag exhibits an almost complete 180° field of view with near-perfect detection rates. This means that users can approach a GazeTag from nearly any angle, including perpendicular to its normal axis (i.e., a user leaning against a wall, looking at a GazeTag attached to the same wall meters away). There is an obvious asymmetry in the accuracy on the vertical axis (Figure 8, right), which is due to the design of the IR receiver we used (which has a slightly asymmetric FoV [61]). As with our previous study, the receptive FoV narrows as distance increases due to the drop-off in transmitted infrared intensity.

6.3 Multi-tag Detection

Results from our multi-tag detection study are plotted in Figure 9. With up to five co-located tags in a small, 25 cm area, performance is strong across all distances: the average detection rate is 99.75% at 4 m, 99.75% at 6 m, and 96% at 8 m. With 6 tags, accuracy drops more significantly to 92% at 4 m, 77% at 6 m, and 41% at 8 m. We optimized our system to reliably support up to five co-located tags (Section 4.4.6), as we felt this represented a reasonable upper bound for the number of tagged objects a user would be facing in many settings. Supporting additional tags is possible by increasing the number of reply repeats and extending the random delay range.

7 Limitations and Future Work

Head gaze is inherently less fine-grained than eye gaze. While prior literature has shown that eye gaze typically deviates from head gaze by no more than approximately 20° [22, 23], this is not always the case. If a user was to rotate only their eyes when glancing at nearby or peripheral objects, GazeTags would not work as reliably. For many use cases, however, we believe head-oriented gaze is a sufficient proxy for eye gaze, and the narrow beam of our head-worn device helps mitigate this issue.

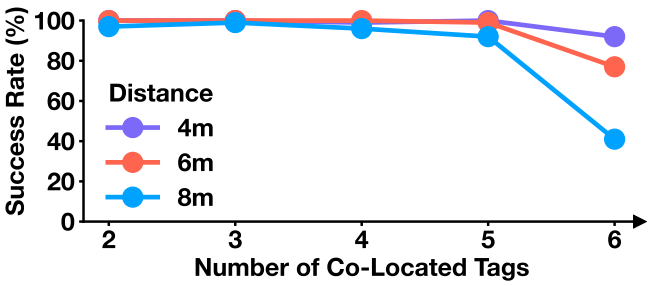


Figure 9: Results of our multi-tag detection study. GazeTag detection rates stay ≥99% with up to 5 co-located tags and at distances up to 6 m.

GazeTags require line-of-sight between the user and the GazeTag in order to function, which could be obstructed through self-occlusion. In our glasses prototype, the hands could block line-of-sight when raised in front of the user, but this is less common when attending to objects meters away. We envision our IR hardware to exist on both earbuds, such that they have a complete FoV when operating together. However, earbuds would still have to contend with long hair, which is an innate limitation of this placement.

In our current implementation, the head-worn device emits a wake-up signal every three seconds, which offers a reasonable latency for many context-aware use cases. GazeTags can also operate in a more event-driven manner, where the head-worn device emits a wake-up signal in response to certain triggers. For instance, an AI assistant could decide to emit a wake-up signal when it needs additional context (e.g., which room is the user in? What objects are they facing?).

As compared to a purely computer-vision-based approach, GazeTags require instrumenting objects in the environment, which could be costly and visually obtrusive. With our current prototype that uses off-the-shelf components, it would cost roughly $220 to instrument 100 objects across a home. Using SMD components and a smaller microcontroller (e.g., ATtiny412) would significantly reduce size. If the cost could be reduced to under $1 and the form factor shrunk to around 1 cm, GazeTags would begin to approach a vision of computing as an infrastructure medium — small, inexpensive, and ubiquitous enough to be embedded into nearly any object [71].

Although power-optimized, GazeTags still require batteries, and the CR2477 coin cell will eventually need replacement. With further power optimizations, GazeTags could operate on a smaller coin cell or run off of harvested solar energy. A mains-powered version of a GazeTag could also enable longer-range (≥20 m) use cases. For instance, buildings and shops could broadcast menus and other information as people look towards them. We note that consumer TV remote controls use the exact same infrared components and have been highly engineered to work outdoors at long range, so while we did not evaluate outdoor performance, we do not anticipate this to be a fundamental limitation.

Finally, head-worn devices could also act as GazeTags. For instance, at a conference, two attendees wearing GazeTag-equipped devices could exchange contact information simply by facing each other, enabling a natural and effortless form of social networking.

8 Conclusion

We presented GazeTags, an inexpensive, low-power, camera-free approach to head-oriented object detection. GazeTags cost roughly $2 each and run for more than a year off a single coin cell battery. Our studies reveal that GazeTags can work well up to a range of 6 m, and with over 99% reliability with up to five co-located tags.

Acknowledgments

We thank Andy Kong for his early work on this project.

References

[1] 3dk Trading GmbH. 2026. PLA Filament IR-Black. https://3dk.berlin/en/special/115-pla-filament-ir-black.html
[2] Adafruit. 2026. ESP8285 [ESP-M2]. https://www.adafruit.com/product/4065

[3] 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 (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 189–196. doi:10.1145/3332165.3347884

[4] Apple. 2026. AirTag. https://www.apple.com/airtag/

[5] Roshan Ayyalasomayajula, Deepak Vasisht, and Dinesh Bharadia. 2018. BLoc: CSI-based accurate localization for BLE tags. In *Proceedings of the 14th International Conference on emerging Networking EXperiments and Technologies (CoNEXT '18)*. Association for Computing Machinery, New York, NY, USA, 126–138. doi:10.1145/3281411.3281428

[6] P. Bahl and V.N. Padmanabhan. 2000. RADAR: an in-building RF-based user location and tracking system. In *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, Vol. 2. IEEE, Tel Aviv, Israel, 775–784. doi:10.1109/INFCOM.2000.832252

[7] Chen Bao, Jiarui Xu, Xiaolong Wang, Abhinav Gupta, and Homanga Bharadhwaj. 2025. HandsOnVLM: Vision-Language Models for Hand-Object Interaction Prediction. *Transactions on Machine Learning Research* (2025). https://openreview.net/forum?id=ehhMFjKnWm

[8] Rahul Bhattacharyya, Christian Floerkemeier, and Sanjay Sarma. 2010. Low-Cost, Ubiquitous RFID-Tag-Antenna-Based Sensing. *Proc. IEEE* 98, 9 (Sept. 2010), 1593–1600. doi:10.1109/JPROC.2010.2051790

[9] Valentina Bianchi, Paolo Ciampolini, and Ilaria De Munari. 2019. RSSI-Based Indoor Localization and Identification for ZigBee Wireless Sensor Networks in Smart Homes. *IEEE Transactions on Instrumentation and Measurement* 68, 2 (Feb. 2019), 566–575. doi:10.1109/TIM.2018.2851675

[10] Roger Boldu, Haimo Zhang, Juan Pablo Forero Cortés, Sachith Muthukumarana, and Suranga Nanayakkara. 2017. InSight: a systematic approach to create dynamic human-controller-interactions. In *Proceedings of the 8th Augmented Human International Conference (AH '17)*. Association for Computing Machinery, New York, NY, USA, 1–5. doi:10.1145/3041164.3041195

[11] Richard A. Bolt. 1980. "Put-that-there": Voice and gesture at the graphics interface. *SIGGRAPH Comput. Graph.* 14, 3 (July 1980), 262–270. doi:10.1145/965105.807503

[12] Michael Buettner, Richa Prasad, Matthai Philipose, and David Wetherall. 2009. Recognizing daily activities with RFID-based sensors. In *Proceedings of the 11th international conference on Ubiquitous computing*. ACM, Orlando Florida USA, 51–60. doi:10.1145/1620545.1620553

[13] Michael Buettner, Richa Prasad, Alanson Sample, Daniel Yeager, Ben Greenstein, Joshua R. Smith, and David Wetherall. 2008. RFID sensor networks with the Intel WISP. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*. ACM, Raleigh NC USA, 393–394. doi:10.1145/1460412.1460468

[14] Nicolas Carion, Laura Gustafson, Yuan-Ting Hu, Shoubhik Debnath, Ronghang Hu, Didac Suris, Chaitanya Ryali, Kalyan Vasudev Alwala, Haitham Khedr, Andrew Huang, Jie Lei, Tengyu Ma, Baishan Guo, Arpit Kalla, Markus Marks, Joseph Greer, Meng Wang, Peize Sun, Roman Rädle, Triantafyllos Afouras, Effrosyni Mavroudi, Katherine Xu, Tsung-Han Wu, Yu Zhou, Liliane Momeni, Rishi Hazra, Shuangrui Ding, Sagar Vaze, Francois Porcher, Feng Li, Siyuan Li, Aishwarya Kamath, Ho Kei Cheng, Piotr Dollár, Nikhila Ravi, Kate Saenko, Pengchuan Zhang, and Christoph Feichtenhofer. 2025. SAM 3: Segment Anything with Concepts. arXiv:2511.16719 [cs.CV] https://arxiv.org/abs/2511.16719

[15] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. 2020. End-to-End Object Detection with Transformers. In *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part I (Lecture Notes in Computer Science, Vol. 12346)*, Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm (Eds.). Springer, 213–229. doi:10.1007/978-3-030-58452-8_13

[16] Yihua Cheng, Haofei Wang, Yiwei Bao, and Feng Lu. 2024. Appearance-Based Gaze Estimation With Deep Learning: A Review and Benchmark. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 46, 12 (Dec. 2024), 7509–7528. doi:10.1109/TPAMI.2024.3393571

[17] Christine Dierk, Molly Jane Pearce Nicholas, and Eric Paulos. 2018. AlterWear: Battery-Free Wearable Displays for Opportunistic Interactions. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, Montreal QC Canada, 1–11. doi:10.1145/3173574.3173794

[18] Estimote. 2026. UWB Tags. https://estimote.com/

[19] Jacqui Fashimpaur, Tovi Grossman, Ben Lafreniere, Naveen Sendhilnathan, Kashyap Todi, Tianyi Wang, Ting Zhang, and Tanya R. Jonker. 2025. Squiggle: Multimodal Lasso Selection in the Real World. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3746059.3747684

[20] Christoph Feichtenhofer, Haoqi Fan, Jitendra Malik, and Kaiming He. 2019. SlowFast Networks for Video Recognition. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*. 6201–6210. doi:10.1109/ICCV.2019.00630

[21] M. Fiala. 2005. ARTag, a fiducial marker system using digital techniques. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2. 590–596. doi:10.1109/CVPR.2005.74

[22] John M. Franchak, Brianna McGee, and Gabrielle Blanch. 2021. Adapting the coordination of eyes and head to differences in task and environment during fully-mobile visual exploration. *PLOS ONE* 16, 8 (Aug. 2021), e0256463. doi:10.1371/journal.pone.0256463

[23] Edward G. Freedman. 2008. Coordination of the eyes and head during visual orienting. *Experimental Brain Research* 190, 4 (Oct. 2008), 369–387. doi:10.1007/s00221-008-1504-8

[24] Yibo Fu, Vivian Shen, Víctor Riera Naranjo, Bolei Deng, Alex Adams, and Josiah Hester. 2025. SoundOff: Low-cost Passive Ultrasound Tags for Non-invasive and Non-Intrusive Smart Home Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 9, 4 (Dec. 2025), 1–32. doi:10.1145/3770666

[25] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (June 2014), 2280–2292. doi:10.1016/j.patcog.2014.01.005

[26] Khusvinder Gill, Shuang-Hua Yang, Fang Yao, and Xin Lu. 2009. A zigbee-based home automation system. *IEEE Transactions on Consumer Electronics* 55, 2 (May 2009), 422–430. doi:10.1109/TCE.2009.5174403

[27] Shubham Goel, Georgios Pavlakos, Jathushan Rajasegaran, Angjoo Kanazawa, and Jitendra Malik. 2023. Humans in 4D: Reconstructing and Tracking Humans with Transformers. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. IEEE, Paris, France, 14737–14748. doi:10.1109/ICCV51070.2023.01358

[28] Google. 2026. Gemini Live. https://gemini.google/overview/gemini-live/

[29] Dan Witzner Hansen and Qiang Ji. 2010. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32, 3 (March 2010), 478–500. doi:10.1109/TPAMI.2009.30

[30] Infrared Data Association. 1997. Serial Infrared Physical Layer Link Specification. https://web.mit.edu/rec/iap/mirror/irda/IrPHY_1_2.PDF.

[31] Tech Insights. 2026. Apple AirTag Teardown. https://www.techinsights.com/blog/apple-airtag-teardown

[32] Arduino IRRemote. 2026. FAST Protocol. https://github.com/Arduino-IRremote/Arduino-IRremote?tab=readme-ov-file#the-fast-protocol

[33] Daehwa Kim, Vimal Mollyn, and Chris Harrison. 2023. WorldPoint: Finger Pointing as a Rapid and Natural Trigger for In-the-Wild Mobile Interactions. *Proc. ACM Hum.-Comput. Interact.* 7, ISS (Nov. 2023), 442:357–442:375. doi:10.1145/3626478

[34] Taejun Kim, Vimal Mollyn, Riku Arakawa, and Chris Harrison. 2026. HiFiGaze: Improving Eye Tracking Accuracy Using Screen Content Knowledge. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3772318.3791339

[35] Rob Knight, Pepijn Kooijmans, Remi Cadene, Simon Alibert, Michel Aractingi, Dana Aubakirova, Adil Zouitine, Russi Martino, Steven Palma, Caroline Pascal, and Thomas Wolf. 2024. Standard Open SO-100 & SO-101 Arms. https://github.com/TheRobotStudio/SO-ARM100

[36] Robert Konrad, Nitish Padmanaban, J. Gabriel Buckmaster, Kevin C. Boyle, and Gordon Wetzstein. 2024. GazeGPT: Augmenting Human Capabilities using Gaze-contingent Contextual AI for Smart Eyewear. https://arxiv.org/abs/2401.17217v2

[37] Adriana Kovashka and Kristen Grauman. 2010. Learning a hierarchy of discriminative space-time neighborhood features for human action recognition. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2046–2053. doi:10.1109/CVPR.2010.5539881 ISSN: 1063-6919.

[38] Kyle Krafka, Aditya Khosla, Petr Kellnhofer, Harini Kannan, Suchendra Bhandarkar, Wojciech Matusik, and Antonio Torralba. 2016. Eye Tracking for Everyone. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2176–2184. doi:10.1109/CVPR.2016.239

[39] Jaewook Lee, Jun Wang, Elizabeth Brown, Liam Chu, Sebastian S. Rodriguez, and Jon E. Froehlich. 2024. GazePointAR: A Context-Aware Multimodal Voice Assistant for Pronoun Disambiguation in Wearable Augmented Reality. 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 408, 20 pages. doi:10.1145/3613904.3642230

[40] Hanchuan Li, Eric Brockmeyer, Elizabeth J. Carter, Josh Fromm, Scott E. Hudson, Shwetak N. Patel, and Alanson Sample. 2016. PaperID: A Technique for Drawing Functional Battery-Free Wireless Interfaces on Paper. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 5885–5896. doi:10.1145/2858036.2858249

[41] Hanchuan Li, Can Ye, and Alanson P. Sample. 2015. IDSense: A Human Object Interaction Detection System Based on Passive UHF RFID. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, Seoul Republic of Korea, 2555–2564. doi:10.1145/2702123.2702178

[42] Life360. 2026. Tile Trackers: Bluetooth Wallet & Key Trackers. https://www.life360.com/tile-trackers

[43] Xin Liu, Silvia L. Pintea, Fatemeh Karimi Nejadasl, Olaf Booij, and Jan C. van Gemert. 2021. No frame left behind: Full Video Action Recognition. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 14887–14896. doi:10.1109/CVPR46437.2021.01465

[44] Sven Mayer, Gierad Laput, and Chris Harrison. 2020. Enhancing Mobile Voice Assistants with WorldGaze. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. doi:10.1145/3313831.3376479

[45] Meta. 2026. Ray-Ban Display Glasses & Neural Band. Specifications, Battery. https://www.meta.com/ai-glasses/meta-ray-ban-display-glasses-and-neural-band/

[46] Microsoft. 2018. Infrared Data Association (IrDA). https://learn.microsoft.com/en-us/previous-versions/windows/desktop/irda/irda-start-page. Documentation overview of the IrDA protocol suite.

[47] Ankit Mohan, Grace Woo, Shinsaku Hiura, Quinn Smithwick, and Ramesh Raskar. 2009. Bokode: imperceptible visual tags for camera based interaction from a distance. In *ACM SIGGRAPH 2009 papers*. ACM, New Orleans Louisiana, 1–8. doi:10.1145/1576246.1531404

[48] Edwin Olson. 2011. AprilTag: A robust and flexible visual fiducial system. In *2011 IEEE International Conference on Robotics and Automation*. 3400–3407. doi:10.1109/ICRA.2011.5979561 ISSN: 1050-4729.

[49] OpenAI. 2026. ChatGPT Voice mode. https://chatgpt.com/features/voice/

[50] OptiTrack. 2026. Active Tag. https://optitrack.com/accessories/active-tag

[51] Preksha Pareek and Ankit Thakkar. 2021. A survey on video-based Human Action Recognition: recent updates, datasets, challenges, and applications. *Artificial Intelligence Rev.* 54, 3 (March 2021), 2259–2322. doi:10.1007/s10462-020-09904-8

[52] Parth H. Pathak, Xiaotao Feng, Pengfei Hu, and Prasant Mohapatra. 2015. Visible Light Communication, Networking, and Sensing: A Survey, Potential and Challenges. *IEEE Communications Surveys & Tutorials* 17, 4 (2015), 2047–2077. doi:10.1109/COMST.2015.2476474

[53] Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, and Jitendra Malik. 2024. Reconstructing Hands in 3D with Transformers. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Seattle, WA, USA, 9826–9836. doi:10.1109/CVPR52733.2024.00938

[54] Rolandos Alexandros Potamias, Jinglei Zhang, Jiankang Deng, and Stefanos Zafeiriou. 2025. WiLoR: End-to-end 3D Hand Localization and Reconstruction in-the-wild. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*. Computer Vision Foundation / IEEE, 12242–12254. doi:10.1109/CVPR52734.2025.01143

[55] Nissanka B. Priyantha, Anit Chakraborty, and Hari Balakrishnan. 2000. The Cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking (MobiCom '00)*. Association for Computing Machinery, New York, NY, USA, 32–43. doi:10.1145/345910.345917

[56] Adria Recasens, Aditya Khosla, Carl Vondrick, and Antonio Torralba. 2015. Where are they looking?. In *Advances in Neural Information Processing Systems*, Vol. 28. Curran Associates, Inc. https://proceedings.neurips.cc/paper/2015/hash/ec8956637a99787bd197eacd77acce5e-Abstract.html

[57] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, Las Vegas, NV, USA, 779–788. doi:10.1109/CVPR.2016.91

[58] A.P. Sample, D.J. Yeager, P.S. Powledge, A.V. Mamishev, and J.R. Smith. 2008. Design of an RFID-Based Battery-Free Programmable Sensing Platform. *IEEE Transactions on Instrumentation and Measurement* 57, 11 (Nov. 2008), 2608–2615. doi:10.1109/TIM.2008.925019

[59] Nordic Semiconductor. 2026. NRF52840-QIAA-R. https://www.digikey.com/en/products/detail/nordic-semiconductor-asa/NRF52840-QIAA-R/7725407

[60] Nordic Semiconductor. 2026. Power Profiler Kit II. https://www.nordicsemi.com/Products/Development-hardware/Power-Profiler-Kit-2

[61] Vishay Semiconductor. 2026. TSOP18338 IR Receiver Module. https://www.vishay.com/docs/82802/tsop181.pdf

[62] Gunnar A. Sigurdsson, Olga Russakovsky, and Abhinav Gupta. 2017. What Actions are Needed for Understanding Human Actions in Videos?. In *2017 IEEE International Conference on Computer Vision (ICCV)*. 2156–2165. doi:10.1109/ICCV.2017.235

[63] Joshua R. Smith, Kenneth P. Fishkin, Bing Jiang, Alexander Mamishev, Matthai Philipose, Adam D. Rea, Sumit Roy, and Kishore Sundara-Rajan. 2005. RFID-based techniques for human-activity detection. *Commun. ACM* 48, 9 (Sept. 2005), 39–44. doi:10.1145/1081992.1082018

[64] Joshua R. Smith, Alanson P. Sample, Pauline S. Powledge, Sumit Roy, and Alexander Mamishev. 2006. A Wirelessly-Powered Platform for Sensing and Computation. In *UbiComp 2006: Ubiquitous Computing*, David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Madhu Sudan, Demetri Terzopoulos, Dough Tygar, Moshe Y. Vardi, Gerhard Weikum, Paul Dourish, and Adrian Friday (Eds.). Vol. 4206. Springer Berlin Heidelberg, Berlin, Heidelberg, 495–506. doi:10.1007/11853565_29 Series Title: Lecture Notes in Computer Science.

[65] J. S. Stahl. 1999. Amplitude of human head movements associated with horizontal saccades. *Experimental Brain Research* 126, 1 (April 1999), 41–54. doi:10.1007/s002210050715

[66] Colin Swindells, Kori M. Inkpen, John C. Dill, and Melanie Tory. 2002. That One There! Pointing to Establish Device Identity. In *Proceedings of the 15th Annual ACM Symposium on User Interface Software and Technology (UIST '02)*. Association for Computing Machinery, New York, NY, USA, 151–160. doi:10.1145/571985.572007

[67] Emmanuel Munguia Tapia, Stephen S. Intille, and Kent Larson. 2004. Activity Recognition in the Home Using Simple and Ubiquitous Sensors. In *Pervasive Computing (Lecture Notes in Computer Science, Vol. 3001)*, Alois Ferscha and Friedemann Mattern (Eds.). Springer, 158–175. doi:10.1007/978-3-540-24646-6_10

[68] Tim van Kasteren, Athanasios Noulas, Gwenn Englebienne, and Ben Kröse. 2008. Accurate Activity Recognition in a Home Setting. In *Proceedings of the 10th International Conference on Ubiquitous Computing (UbiComp '08)*. Association for Computing Machinery, New York, NY, USA, 1–9. doi:10.1145/1409635.1409637

[69] Roy Want, Andy Hopper, Veronica Falcão, and Jonathan Gibbons. 1992. The active badge location system. *ACM Trans. Inf. Syst.* 10, 1 (Jan. 1992), 91–102. doi:10.1145/128756.128759

[70] A. Ward, A. Jones, and A. Hopper. 1997. A new location technique for the active office. *IEEE Personal Communications* 4, 5 (Oct. 1997), 42–47. doi:10.1109/98.626982

[71] B. Warneke, M. Last, B. Liebowitz, and K.S.J. Pister. 2001. Smart Dust: communicating with a cubic-millimeter computer. *Computer* 34, 1 (2001), 44–51. doi:10.1109/2.895117

[72] 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 (UIST '19)*. Association for Computing Machinery, New York, NY, USA, 175–187. doi:10.1145/3332165.3347954

[73] Xiaoying Yang, Jacob Sayono, Jess Xu, Jiahao Nick Li, Josiah Hester, and Yang Zhang. 2022. MiniKers: Interaction-Powered Smart Environment Automation. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 3 (Sept. 2022), 1–22. doi:10.1145/3550287

[74] Xiaoying Yang, Jacob Sayono, and Yang Zhang. 2023. CubeSense++: Smart Environment Sensing with Interaction-Powered Corner Reflector Mechanisms. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. ACM, San Francisco CA USA, 1–12. doi:10.1145/3586183.3606744

[75] Huizhong Ye, Chi-Jung Lee, Te-Yen Wu, Xing-Dong Yang, Bing-Yu Chen, and Rong-Hao Liang. 2022. Body-Centric NFC: Body-Centric Interaction with NFC Devices Through Near-Field Enabled Clothing. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (DIS '22)*. Association for Computing Machinery, New York, NY, USA, 1626–1639. doi:10.1145/3532106.3534569

[76] Moustafa Youssef and Ashok Agrawala. 2005. The Horus WLAN location determination system. In *Proceedings of the 3rd international conference on Mobile systems, applications, and services (MobiSys '05)*. Association for Computing Machinery, New York, NY, USA, 205–218. doi:10.1145/1067170.1067193

[77] Ben Zhang, Yu-Hsiang Chen, Claire Tuna, Achal Dave, Yang Li, Edward A. Lee, and Björn Hartmann. 2014. HOBS: Head Orientation-Based Selection in Physical Spaces. In *Proceedings of the 2nd ACM Symposium on Spatial User Interaction (SUI '14)*. Association for Computing Machinery, New York, NY, USA, 17–25. doi:10.1145/2659766.2659773

[78] Yang Zhang, Yasha Iravantchi, Haojian Jin, Swarun Kumar, and Chris Harrison. 2019. Sozu: Self-Powered Radio Tags for Building-Scale Activity Sensing. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New Orleans LA USA, 973–985. doi:10.1145/3332165.3347952

[79] Minghui Zhao, Tyler Chang, Aditya Arun, Roshan Ayyalasomayajula, Chi Zhang, and Dinesh Bharadia. 2021. ULoc: Low-Power, Scalable and cm-Accurate UWB-Tag Localization and Tracking for Indoor Applications. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 5, 3 (Sept. 2021), 140:1–140:31. doi:10.1145/3478124