TangiAR: Markerless Tangible Input for Immersive Augmented Reality with Everyday Objects

Neil Xu Fan
University of British Columbia
Vancouver, BC, Canada
fanxu104@cs.ubc.ca

Xincheng Huang
University of British Columbia
Vancouver, BC, Canada
xincheng.huang@ubc.ca

Robert Xiao
University of British Columbia
Vancouver, BC, Canada
brx@cs.ubc.ca

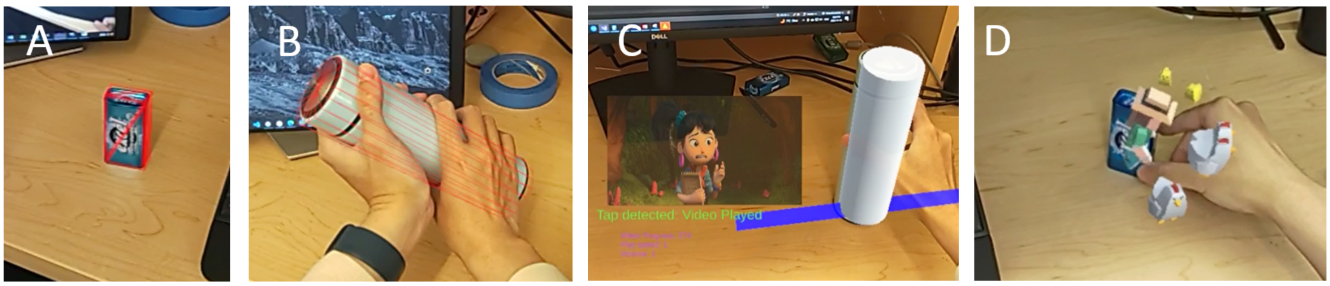


Figure 1: *TangiAR* enables tangible interactions using everyday objects in immersive augmented reality. (A) *TangiAR* tracks the 3D position and rotation of objects without the need for additional visual markers or texture registration, (B) even under substantial occlusion. We showcase the performance of the system with two example applications: (C) a tangible controller for an AR video player; and (D) an AR game: Farm, where the object is used as a tangible proxy.

Abstract

Tangible interactions with everyday objects have been shown to be fast, accurate, and natural, and have shown promise when combined with immersive augmented reality. However, implementing tangible controls presents considerable challenges. Previous works in the field either rely on additional tracking markers on objects, inadvertently shifting the difficulty to users, or are too computationally demanding for real-time operation on a head-mounted display (HMD). We propose *TangiAR*, a tangible control system which tracks everyday objects without the need for fiducial trackers, enabling them as passive controllers and virtual proxies in AR applications. *TangiAR* additionally enables hand and finger proximity interactions with tangibles, further expanding the interaction space. *TangiAR* can run on an unmodified Microsoft HoloLens 2, making it immediately practical. We evaluated the performance of *TangiAR* through a technical evaluation, including occlusion robustness and tracking accuracy tests, and a user study which examined the usability of our markerless object tracking system in various AR interactions.

CCS Concepts

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

VRST '25, Montreal, QC, Canada

ACM ISBN 979-8-4007-2118-2/2025/11
https://doi.org/10.1145/3756884.3766028

ACM Reference Format:
Neil Xu Fan, Xincheng Huang, and Robert Xiao. 2025. *TangiAR*: Markerless Tangible Input for Immersive Augmented Reality with Everyday Objects. In *31st ACM Symposium on Virtual Reality Software and Technology (VRST '25), November 12–14, 2025, Montreal, QC, Canada.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3756884.3766028

1 Introduction

Immersive augmented reality (AR) seamlessly overlay virtual content onto the physical world. While traditional AR input techniques, such as gesture, gaze, and voice, are expressive and natural, but they fall short in precision and lack haptic feedback. Physical controllers, on the other hand, offer both precision and tactility, but requiring users to carry a separate device undermines the convenience and fluidity expected from AR systems.

Tangible input [14, 28, 30, 38] presents a compelling alternative. By turning everyday objects into adaptable input devices, tangibles align closely with the spatial and embodied nature of AR. They provide inherent, low-cost tactile feedback and support intuitive, familiar interactions. Prior research also shows that tangible controls can match the precision of mouse and touch input, while achieving faster task completion and reducing user workload [5, 26]. Researchers have also highlighted the value of creating ad hoc tangibles [14], which turns mundane objects into digital interfaces. Such ad hoc tangibles often require no additional instrumentation, making tangible interaction more scalable and seamless.

However, building a tangible control system with daily objects remains challenging due to the computational demands of real-time object tracking and the need for additional tracking infrastructure such as markers [11, 15] or external cameras [54]. Few commercial products embrace this approach, and most existing systems, whether research or commercial, rely on tracking infrastructure such as markers [11, 15], external cameras [54], or specialized surfaces [38] to function reliably. These requirements impose barriers for everyday users: instrumenting environments is costly, and creating or acquiring marked objects adds friction to adoption. This brings us to our research question: How can we provide markerless, minimal-setup, and occlusion-robust tracking for tangible AR interaction?

This paper presents *TangiAR*, a novel markerless tangible input system for augmented reality on head-mounted displays (HMDs). Our work advances the field of tangible AR interaction—particularly for HMD-based systems—by tackling key technical challenges and offering empirical evidence of the system's effectiveness. The main research contributions are: (1) a markerless tangible AR system running on unmodified HoloLens 2, (2) technical evaluation of tracking performance under realistic conditions, (3) user study demonstrating usability benefits, and (4) open-source release at https://github.com/UBC-X-Lab/TangiAR.git.

2 Related Work

There is a long history of using tangible objects to interact with AR systems. This section provides an overview of the most relevant works to ours, followed by an overview of object tracking methods.

2.1 Tangible Augmented Reality

Tangible interactions, as introduced by Ishii et al. in Tangible Bits[30], aim to bridge the gap between the physical and digital worlds. Virtual data is coupled with physical objects, allowing users to interact with digital content by manipulating the real world. This concept extended to AR as Tangible AR [7], where surfaces and everyday objects become 2D touch interfaces [22, 25, 42, 57].

Prior work has explored surface-top touch input with occlusion detection [34] and depth sensing [18, 56, 58, 59]. Individual physical objects also offer a rich channel for AR interaction. Poupyrev et al. [44] used gestures like picking up, tilting, or releasing objects to trigger commands. Others mapped object semantics to digital functionality for easier learnability [4, 6, 39]. Inter-object interaction, such as events triggered by relative positions or orientations, has also been explored [8]. Most systems use printed markers for object tracking, enabling real-time pose estimation.

Interacting with 3D environments via tangible objects feels natural to users and yields measurable performance benefits. Besancon et al. [5] showed tangible controls match mouse/touch precision while reducing task time and workload. Bozgeyikli et al.[9] found users preferred tangible proxies over controllers and gestures. Dünser [17] reported faster performance using tangible sliders over paddles. Hettiarachchi and Wigdor [26] used everyday objects as tactile proxies. Greenslade et al. [23] found no consensus on object preference, suggesting users should choose their own proxies. These findings informed our design.

Various methods exist for tangible tracking. iaTAR [35] used cube-shaped props with visual markers on each face. PAIR [51] combined IMU data with screen-displayed markers to turn phones into 6-DOF controllers. To improve aesthetics, some systems [11, 32] used custom markers with distinctive features. Others enhanced accuracy with IR markers [54], embedded LEDs [48], or passive reflective materials [2]. Prior research has also combined with motion capture methods such as IMUs [43, 47]. Alternative sensing approaches use magnetics [38], radar [61], bio-sensing [53], or spatial touch fields [46].

Besides the above systems that may require complex setups, researchers have explored more accessible alternatives. iCon [11] let users attach visual stickers to household objects. Drogemuller et al.[15] proposed QR bands that wrap around objects on demand. Walsh et al.[54] studied everyday object interactions, while depending on an external tracking system such as Vicon [1]. Tangi [19] introduced a toolkit for rapidly building physical proxies. Some systems go further, enabling ad-hoc tangible control. Instant User Interfaces[13] used depth cameras for marker-free tracking with touch detection, though the GPU requirements were too demanding for AR HMDs. Du et al.[16] used UV features, and Funk [21] combined feature detection with depth sensing. However, neither of them reported system performance in the paper.

2.2 3D Object Tracking

As classified by Chen [10], 3D object tracking algorithms fall into three categories: generative, discriminative, and deep learning-based. Generative models estimate the most probable pose by searching near the previous state. With robust keypoint detectors like SIFT [41] and SURF [3], many rely on template matching [33, 40, 52], minimizing feature-to-feature distance. These methods depend on distinctive visual features of target objects.

Discriminative trackers classify image regions into foreground (object) and background, then infer pose from the foreground. Early examples include PWP3D [45]. Li et al.[37] used SVMs to adapt to background variation, while other works[12, 27] enhanced foreground-background separation. SRT3D [49] introduced an efficient tracking formulation, and ICG [50] integrated depth data to improve performance.

With the rise of deep learning, many trackers now use neural networks to predict pose changes from image input. Approaches include Siamese [36, 62], RNN [60], and attention-based architectures [55]. These methods offer high robustness and accuracy but are computationally intensive and require GPU acceleration.

3 Implementation

Existing tangible systems either rely on tracking markers—which are cumbersome and visually intrusive—or object detection pipelines that are too computationally intensive for HMDs. *TangiAR* offers an efficient, markerless alternative. Our primary standalone prototype runs entirely on the HoloLens using its forward camera (Figure 2a). To address its limited field of view, we also built a second prototype with a wide-FOV camera and offloaded tracking to a PC (Figure 2b).

3.1 Tracking Engine

Our tracking engine builds on the Iterative Corresponding Geometry (ICG) algorithm [50], a discriminative tracker that segments foreground and background, then estimates object pose by aligning the detected boundary with pre-computed object views. These views are generated offline from different angles, allowing efficient

[1]https://www.vicon.com/

(a) The standalone setup

(b) The wide-field-of-view setup

Figure 2: The hardware setups

real-time optimization via rigid body transforms to produce a 6-DOF pose.

We chose ICG for its real-time performance and robustness to occlusion—critical for tangible controls involving hand interaction. Unlike keypoint-based methods (e.g., SIFT [41], SURF [3]), ICG relies on boundary features and thus does not require object textures, making it well-suited for featureless everyday objects. Its tolerance for imperfect geometry further supports our use case. However, ICG lacks object detection and requires an initial pose. Our system addresses this via manual alignment and implicit tracking when the object leaves the view, detailed in Sections 3.5 and 3.7.

3.2 Standalone Prototype

After selecting the tracking algorithm, we integrated it into a tangible control system. To demonstrate feasibility, we built a self-contained prototype running entirely on HoloLens 2, using only its onboard color camera to save computational resources. We exposed ICG's core functions—uploading camera intrinsics, updating images, running tracking, and exporting poses—as Unity/C# APIs, and compiled the tracking engine as a shared library.

To maintain real-time performance while allowing application logic to run smoothly, we adopted a multi-threaded architecture. The main UI thread is reserved for the app, while separate threads handle image capture, pixel conversion, and pose estimation. Each tracked object runs on its own thread. We capture images at 896×504 resolution, balancing quality and performance. Our prototype supports tracking up to two objects at 60 fps. Future work could adopt a dedicated multi-object tracker for improved efficiency.

Computed poses are expressed as a position vector and a rotation quaternion. To reduce jitter, we apply an exponentially-weighted moving average (EWMA) to position, and SLerp with EWMA to rotation. Position changes over the last 10 frames are stored in a circular buffer to support velocity-based interaction.

3.3 Wide-Field-of-View Prototype

Our standalone prototype demonstrates a fully self-contained, markerless tangible system on HoloLens 2. While the HoloLens includes four wide-angle infrared cameras for spatial tracking, these are not accessible to user programs. Its front-facing color camera has a narrow field of view (FOV), limiting tangible interaction. To address this, our second prototype mounts an Azure Kinect—with a wide-FOV color camera and depth sensing—on top of the HoloLens visor (Figure 2b). The Kinect connects via USB-C to a PC (i7-11700 @ 2.5 GHz, 64GB RAM, Windows 10), which runs the tracking engine and streams object poses to the HoloLens over WiFi via UDP. Performance tests (Sections 4.1, 4.2, and 4.3) are based on this setup. This wide-FOV prototype illustrates *TangiAR*'s full potential and anticipates future headsets integrating similar cameras for enhanced interaction.

3.4 Model Generation

Our tracking engine requires a 3D model of the object (but not its texture) to function. The model is preprocessed to extract boundary information from multiple angles, accelerating pose estimation. However, accurate modeling is time-consuming and technically challenging. One solution is 3D scanning [24], which can be done with standard depth cameras [31]. In our prototype, we instead approximate objects using geometric primitives. Many everyday objects can be roughly modeled this way, and our evaluation examines how the tracker handles such imperfections. For complex shapes, 3D scanning remains a viable alternative.

To register a new object, our tool lets users select a primitive (sphere, cuboid, or cylinder) and input relevant dimensions (e.g., width, height, radius). It then uses OpenSCAD [1] to generate the mesh, preprocesses it for tracking, and saves the resulting model.

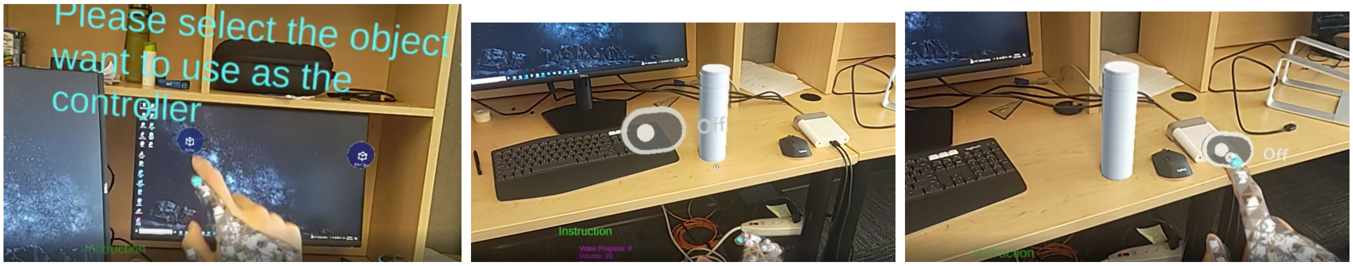
3.5 Object Choice and Initial Alignment

To use tangible objects as controllers, the system maps their 6-DOF pose—translation and rotation—to application inputs. However, due to symmetry, some geometries lack certain degrees of freedom. For instance, a cylinder has ambiguous rotation along its vertical axis, while a sphere lacks meaningful rotation entirely. The system accounts for such limitations by recording the available degrees of freedom during object registration based on the selected geometric template.

Application designers can specify which rotational DOFs are required for a given control. For example, a translation-only control may need no rotation, while a tilt-based input might require two. At runtime, the system filters registered objects based on these requirements and prompts the user to select a compatible one. Once selected, the user spawns the virtual object at a desired location. A mesh of the object is rendered, and the user is prompted to align the real object with it to establish the initial pose. Tracking begins after the user confirms alignment with a toggle (Figure 3).

3.6 System Calibration

Object poses from our tracking engine are reported in camera coordinates and must be transformed into the virtual environment's space. In the wide-FOV prototype, the camera pose is estimated via a calibrated rigid-body transform using an affine matrix (described below). In the standalone version, the Windows Holographic API directly provides the camera pose. We combine this with the tracking output to compute object poses in Unity coordinates. The API supplies a per-frame camera-to-world ($C2W$) matrix. Due to differing coordinate conventions, we apply a sequence of basis transformations: $T2S$ (180° x-rotation), $S2U$ (reflection across the x–y plane),



(a) Step 1: select an object **(b) Step 2: spawn virtual object by pinching** **(c) Step 3: click toggle to start tracking**

Figure 3: The process of initial alignment

and their inverses $U2S$, $S2T$. The tracking engine's pose M is converted into Unity space as M', as shown in Equation 2.

We observed a consistent positional offset between the tracked pose and the rendered object in Unity, linearly related to the object's camera coordinates. To correct this, we defined an affine offset matrix. This correction models the offset between the physical camera and HoloLens coordinates as a rigid-body transform. For the wide-FOV prototype, we aligned the Azure Kinect with the HoloLens and fine-tuned the transform by rendering an asymmetric cuboid in Unity and visually aligning it to the real object.

In the standalone version, where direct visual alignment wasn't possible, we built a Unity calibration tool. Users adjusted a rendered object's position via sliders until it overlapped the physical one, recording 25 (camera pose, offset) pairs. We then used RANSAC to fit an affine matrix, which significantly improved alignment in subsequent tracking.

$$M = \begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}, \quad \text{Offset} = \begin{bmatrix} 0(3*3) & A*T \\ 0(3*1) & 0 \end{bmatrix}; \tag{1}$$

$$M' = S2U * C2W * I2S * (M + \text{Offset}) * S2I * U2S. \tag{2}$$

3.7 Handling Out-of-View Objects

Visual marker-based tangible systems detect markers in the camera feed to compute object pose. In contrast, iterative optimization methods estimate pose changes frame-to-frame, eliminating the need for markers. However, these methods cannot recover tracking once the object leaves the camera's view, even if it re-enters later. This becomes a practical issue in immersive AR, where users frequently move their heads, especially with the narrow field-of-view (FOV) camera on HoloLens 2.

To address this, we track object pose in camera coordinates per frame. If the object is about to exit the view, we record its world pose and pause tracking, while continuing to collect camera orientation. Once the object is expected to re-enter the view, *TangiAR* reprojects its pose into camera space to resume tracking. This frees users from keeping objects continuously visible but assumes the object remains stationary while out of view. We revisit this assumption in the Limitations section.

Since objects can still be visible through the see-through display even when outside the tracking range, we implemented a color-coded warning system (Figure 4). As the object nears the camera boundary, it turns yellow; once out of range, it turns red and tracking pauses, thus signaling the user to stop moving the object. This feature addresses the limited FOV of the HoloLens camera. In contrast, our wide-FOV prototype, which uses a broader-view camera, reduces the need for such warnings.

3.8 Touchable Objects

To enhance interactivity, we added a touchable object feature to *TangiAR* using the HoloLens's built-in hand tracking and the known geometry of the tracked object. Touch is detected by measuring the distance between the fingertip joint and the object surface; a threshold of 1.5 cm is used to trigger a touch event. The system supports different actions based on finger, hand, or contact location, allowing users to interact with the object in varied ways.

4 Technical Evaluation

To evaluate *TangiAR*'s performance, we conducted experiments using the wide-FOV hardware prototype to address three key questions: 1) How does *TangiAR* perform with imperfect or complex object models? 2) How robust is it under hand occlusions? 3) What is the system's latency, and what are its sources?

4.1 Model Error Tolerance

Using the tracking engine requires a 3D model of the target object. However, asking non-technical users to model objects precisely can be challenging. Our prototype offers a shortcut by letting users approximate objects with common geometric primitives, though these often differ from the real object. Such modeling errors are expected to impact tracking accuracy. To evaluate *TangiAR*'s tolerance to these discrepancies, we conducted an experiment measuring the effect of model–object mismatch.

To control variables, we modeled five everyday objects using the same cylinder template (Figure 5): 1) a cylindrical water bottle, 2) a roll of tape, 3) a coffee mug, 4) a narrow-neck bottle, and 5) a smooth rock. While our system supports other geometries, these objects vary in how closely they match a true cylinder and are representative of common household items. For the mug and bottle, the cylinder was defined to fit only the main body, excluding the handle and neck.

For each object, we recorded the 6-DOF pose reported by *TangiAR* over 30 seconds while keeping the object still and moving the headset. This tests dynamic tracking stability, as object motion within the camera view causes continuous pose updates. In world

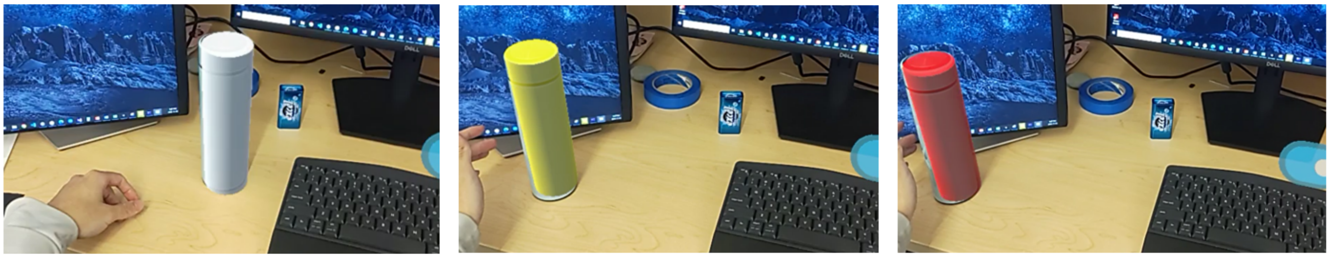


Figure 4: Color-coded warning system: the object turns yellow when it is moved close to the boundary of the camera to warn the users. It turns red when it is out of the camera frame and the tracking is paused.

coordinates, the object should remain stationary, so any observed movement reflects tracking error.

We quantify tracking stability using Mean Absolute Deviation (MAD) and Standard Deviation (SD), which reflect the noise level for the tracking of each object. Raw pose data is visualized in the supplemental material. Each cell in Table 1 reports the average measurement across all dimensions. These results demonstrate that TangiAR performs robustly with imperfect object models, maintaining acceptable tracking accuracy even when geometric primitives approximate complex real-world objects.

4.2 Hand Occlusion Tolerance

To evaluate hand occlusion's impact on tracking performance, we tested 3 differently sized objects: a mint box (cube model, 2×4×7 cm), a roll of tape (cylinder model, $R = 6$ cm, $h = 2$ cm), and a large water bottle (cylinder model, $R = 3.5$ cm, $H = 23$ cm), as shown in Fig. 5. We selected 2 common grasp gestures per object from [20], representing 15% and 50% occlusion levels (Fig. 6). Following the same procedure as before, each object was fixed in place and held with the specified grasp. We recorded *TangiAR*'s 6-DOF output while moving the headset to vary the camera angle, with the noise levels summarized in Table 2. TangiAR shows good robustness to hand occlusion, maintaining usable tracking performance up to 50% occlusion levels, particularly for larger objects.

4.3 Latency

Low latency is critical for delivering an immersive real-time experience. *TangiAR* exhibits a Mixed Reality Capture (MRC)-measured latency of 130–160 ms, based on the number of frames it takes for a virtual object to catch up with a fast-moving real one. This latency has four main components: camera capture, tracking computation, network communication, and rendering.

Camera capture time averages 120–140 ms, measured by comparing real-world movement to Azure Kinect video output using a 120 fps phone camera. Tracking and network latency are measured via timestamps, showing 1–3 ms and 2–3 ms, respectively. HoloLens rendering latency is difficult to measure directly, but assuming a 60 fps display, we estimate it at 8 ms.

In the standalone version, latency includes front camera capture, tracking, and rendering. Tracking on the HoloLens CPU takes 11–16 ms per frame. However, the MRC-measured latency exceeds 250 ms. Possible causes are discussed in Section 8.3.

The system achieves acceptable latency (130-160ms) for real-time interaction, with camera capture being the primary bottleneck that could be improved in future AR hardware.

5 Example Applications

We developed two applications to demonstrate TangiAR's versatility across different interaction paradigms. The augmented video player represents precision-focused tasks requiring fine-grained control (progress and volume adjustment), while the immersive sandbox game demonstrates spatial manipulation and embodied interaction where the tangible object serves as a virtual proxy. These scenarios cover the primary use cases for tangible AR: precise control interfaces and immersive spatial interaction.

5.1 Augmented Video Player

Our *TangiAR* video player leverages a tangible object as a controller, eliminating the need for visible on-screen controls. This approach addresses the common interface design challenge of balancing functionality with usability. Users can:

(1) Move the object along x/y axes to control video progress and volume
(2) Tilt the object to adjust playback speed and navigate between videos
(3) Utilize touch inputs for play/pause functionality and accessing a context menu

We implemented a state machine to effectively decouple inputs, engaging specific modes based on predefined movement thresholds. Rotation inputs are managed by repositioning the model's origin and employing a 3D vector representation. The system also allows for clutching, enabling users to reposition the object as needed.

5.2 Immersive Sandbox Game

We designed "Farm", an interactive game that utilizes a tangible object as a spatial anchor for a virtual farmer character. This approach aims to enhance immersion and ease of use by mapping virtual elements onto familiar, tactile real-world objects. The tangible's orientation and velocity drive the farmer's movements and animations. Users can interact with in-game elements, such as chickens, and perform actions like saluting and feeding. This application showcases how *TangiAR* can be employed to create engaging and tangible mixed reality experiences.

(a) Cylindrical bottle (b) Tape (c) Coffee mug (d) Narrow-neck bottle (e) River rock

Figure 5: Five objects used to test model-object error tolerance. We only use cylinder models here for consistency.

Objects	Positional MAD (cm)	SD (cm)	Rotational MAD (deg)	SD (deg)
Cylindrical bottle	0.044	0.053	0.178	0.221
Tape	0.109	0.125	0.394	0.489
Coffee mug	0.049	0.061	0.212	0.269
Narrow-neck bottle	0.078	0.097	0.422	0.524
River rock	0.027	0.035	0.412	0.517

Table 1: Model error test noise level

	0% Occlusion				15% Occlusion				50% Occlusion			
	Positional (cm)		Rotational (deg)		Positional		Rotational		Positional		Rotational	
	MAD	SD	MAD	SD	MAD	SD	MAD	SD	MAD	SD	MAD	SD
Mint Box (small)	0.031	0.034	0.459	0.545	0.067	0.095	0.536	0.660	0.072	0.092	0.528	0.686
Tape (medium)	0.041	0.051	0.273	0.341	0.025	0.032	0.243	0.306	0.071	0.092	1.199	1.418
Bottle (large)	0.034	0.043	0.129	0.163	0.047	0.058	0.123	0.169	0.122	0.145	0.471	0.568

Table 2: Occlusion test noise level

6 User Study

We conducted a user study to evaluate *TangiAR*'s usability, involving 10 participants (8 male, ages 23-31, mean 26). Participants were compensated $24 for a 90-minute session. 2 participants self-identified as VR/AR experts, while the rest had minimal AR experience. Our study was approved by our institution's Behavioral Research Ethics Board (BREB). We adapted the application scenarios described from the previous section for this user study:

(1) **Video Player Comparison.** Participants used three interfaces: (1) *TangiAR* (see Figure 7A and 7B), (2) YouTube in web browser, and (3) HoloLens built-in player (a floating 2D video player) to perform common video control actions. We measured task completion times and collected NASA-TLX workload assessments. Post-task interviews gathered qualitative feedback on each interface. Note that the NASA-TLX scores were collected using the standard 21-point scale and converted to a 0-9 scale for analysis and presentation.

(2) **Immersive Game Exploration.**: Users engaged with the Farm game, utilizing a tangible object as a proxy for their virtual character. This open-ended task allowed for observation of user behavior and collection of subjective experiential feedback.

All tasks were performed using the standalone *TangiAR* prototype. The video player interfaces were presented in randomized order to mitigate learning effects. We selected YouTube and HoloLens players for comparison due to their contrasting designs in functionality and usability. The study aimed to compare *TangiAR*'s speed, accuracy, and usability against existing AR interfaces, while also gathering insights on its potential for creating immersive experiences. Our study was approved by our institution's ethics board.

7 Results

7.1 Video Player Study

First, we compared task completion time for coarse browsing and fine adjustment across 3 video players. Action mappings for each

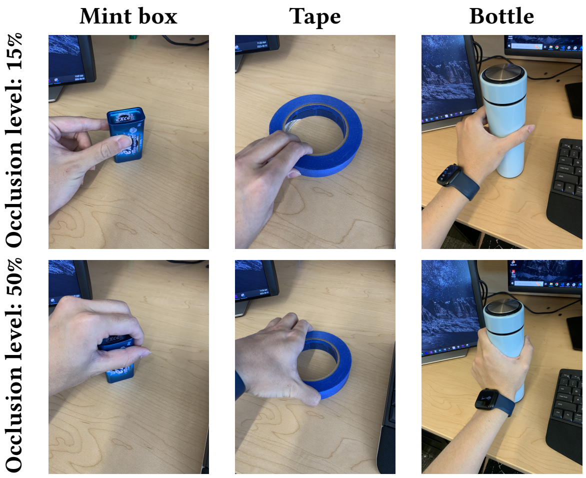


Figure 6: Grasp gestures used to test occlusion tolerance

player are listed in Table 3. Completion times were analyzed using Kruskal-Wallis tests due to the non-normal distribution typical of reaction time data. Significant differences were found for tasks available across all three systems: coarse adjustment ($H(2) = 13.905$, $p < 0.001$), fine adjustment ($H(2) = 4.015$, $p < 0.05$), and play/pause actions ($H(2) = 20.505$, $p < 0.001$). For tasks with limited system availability: slide/tilt actions comparing TangiAR vs HoloLens ($H(1) = 5.851$, $p < 0.05$) and volume control comparing TangiAR vs YouTube ($H(1) = 9.882$, $p < 0.01$). Some actions were excluded due to added latency or multi-step triggers—for example, YouTube's previous/next and speed controls require navigating menus or are affected by network delay.

Figure 8a presents task completion times (in seconds) and user-reported workload. Each workload dimension showed a statistically significant difference between the three systems: Mental Demand ($F(2, 27) = 8.899$, $p < 0.01$), Physical Demand ($F(2, 27) = 15.947$, $p < 0.001$), Temporal Demand ($F(2, 27) = 5.678$, $p < 0.01$), Performance ($F(2, 27) = 18.587$, $p < 0.001$), Effort ($F(2, 27) = 35.559$, $p < 0.001$), and Frustration ($F(2, 27) = 32.761$, $p < 0.001$).

Interview responses showed unanimous preference for the *TangiAR* control system. Common reasons included: "It is more accurate", "Very easy to learn", "Holding something gives you more control," and "Left-hand touch interactions are easy and responsive." When asked about drawbacks, users were similarly aligned: "If I block the object with my wrist, it loses track," and "When the object moves out of view, tracking stops—I have to follow it with my head." One participant rated *TangiAR* highest in Mental Demand due to this, while others noted it as a limitation but found it acceptable.

Some opinions were more divided. Participants P3, P4, and P6 praised using object rotation as a binary input, with P4 likening it to a joystick: "It is very responsive." In contrast, P10 found it less intuitive, preferring to twist the controller's tail instead.

For the YouTube player, all users experienced frequent mistriggers—accidentally hitting play/pause, volume, or opening the contextual menu. This slowed task performance and increased frustration and workload. Many attributed the issue to small button size. The HoloLens native player performed better in this regard; four users preferred it over YouTube, while three felt the difference was minimal, citing only minor UI changes.

7.2 Immersive Game: "Farm"

User feedback on the simple game was overwhelmingly positive. Many described it as "very smooth" and "a unique experience." Despite minimal instruction—just "go ahead and explore"—all users navigated the game with ease. Having already learned the system during the video player task, they independently selected a controller object, spawned its virtual counterpart, and aligned it with the real one. Once tracking began and the farmer avatar appeared, users naturally grabbed the object and began moving it.

All participants discovered the walk/run mechanic on their own. If a user kept moving the farmer for over a minute, the researcher would prompt: "Try a jump." In response, everyone instinctively lifted the object to trigger a jump—one user even flipped it and asked, "Can't he do a backflip?" When left-hand touch interactions weren't discovered organically, the researcher would later introduce them, allowing users to feed and chase the chickens.

From the interview after the game, we collected some insightful user feedback:

> **P3:** "*Although I didn't discover the left-hand actions by myself, when you told me it, I felt very natural the small farmer should do something when I poke him with my finger. It feels like poking a real person*"

> **P4:** "*It is a very unique experience. It is hard to compare it with a traditional computer game. Because each of them has its own area. I guess some computer games can't be replaced by this, like an intense first-person shooting game. For some other games, like a strategy war simulator, this feels like the best interaction.*"

> **P2:** 11 *When I interacted with virtual objects in HoloLens, pinching my fingers together to grab the object gives me a feeling of breaking the immersive experience. It feels like the 3D experience is back to 2D again. But with this tangible object. I didn't have such a feeling. I forgot about the control in the game.*"

Regarding downsides, one user noted that moving the object too quickly caused tracking loss, which broke the immersive experience. Another mentioned that mismatched object weight could affect realism—if the physical controller feels too different from expectation, it may disrupt immersion. Other suggestions focused more on game complexity than on the tangible control itself.

User feedback deepened our understanding of the system. While this study combined quantitative task timing with qualitative impressions, the visible joy on participants' faces while guiding the farmer goes beyond what can be captured in text. We encourage viewing the video figure to better appreciate the experience. We see *TangiAR* as a promising platform for immersive storytelling in future AR systems.

Figure 7: A demonstration of the user study setups: (A) a y-directional translation movement of the controller object adjusts the volume of the video player. A vertical virtual bar is rendered to indicate the control range. (B) an x-directional movement controls the play head. A horizontal bar indicates the progress bar. (C) The AR game: Farm, where the tangible object is used as a physical proxy of the game character.

	TangiAR	HoloLens Player	YouTube
playhead	translation in x	drag/poke the progress bar	drag/poke the progress bar
volume control	translation in y	N/A (uses system volume)	drag/poke the volume bar
pause/play	left index finger tap	tap the button	tap the button/video
next/previous video	tilt in x	slide or press the buttons	press the buttons (not evaluated)
speed up/down	tilt in y	N/A (not available)	press buttons in settings (not evaluated)
context menu	left thumb tap	long press (3s; not evaluated)	long press (1s)

Table 3: Input mappings across video players

8 Discussion

8.1 Model Error Tolerance

In the model error tolerance experiment, the baseline for positional output is the mean reported position, as the object remains stationary. For rotation, assuming a level surface, the baseline is 0 degrees. Variance from these baselines reflects the system's precision in estimating spatial orientation. For tangible controls, the key metric is noise level, indicating the controller's error margin.

The tracking system showed strong performance across objects of varying complexity. The tape, with a hollow core and cylindrical boundary, exhibited slightly higher noise (up to 0.5 cm and 5°) than ideal models like the cylindrical bottle. Still, noise remained within acceptable bounds. The coffee mug, despite its handle, showed low noise—likely because the tracker focuses on the main body, reducing the handle's impact. In contrast, the narrow-neck bottle posed more challenges due to its larger proportion of complex geometry.

These results highlight the system's ability to prioritize simpler, trackable regions within more complex objects—a practical strength for real-world use, even if it introduces minor tracking noise. The river rock, despite its non-cylindrical shape, further illustrates this flexibility, though its irregular form led to positional offsets that limit its utility as a precise controller.

8.2 Hand Occlusion Tolerance

Results from the occlusion tolerance test show that hand occlusion impacts rotation more than position. Larger objects remain more stable, as the system can better estimate pose when more object pixels are visible. Occlusion affects rotation in two ways: by shifting the average and increasing noise. At 15% occlusion, *TangiAR*'s performance remains largely unaffected, but at 50%, both rotational drift and noise increase noticeably. Still, for applications that don't require precise rotation, the system remains usable under 50% occlusion—especially when tracking larger objects.

Compared to marker-based systems that fail completely when markers are occluded, our boundary-based approach degrades gracefully, maintaining approximate pose even under partial occlusion. This represents a practical advantage for natural hand-object interaction, though users must still be mindful of grip placement for rotation-sensitive tasks.

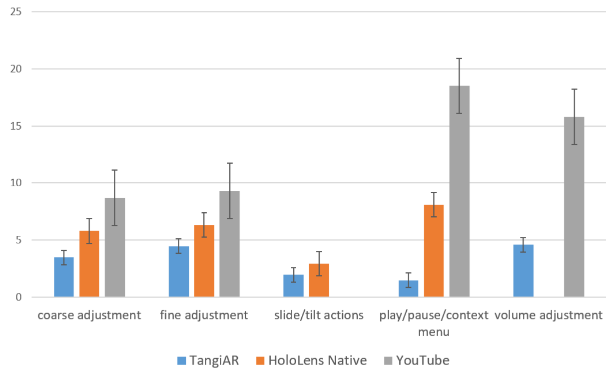
8.3 Latency

The current system exhibits a total latency of 130–160 ms, with potential for improvement. Future AR headsets with lower-latency cameras and faster CPUs could reduce this below 30 ms. While the standalone version shows higher measured latency than the wide-FOV prototype (see Video Figure), the difference is not perceptible in practice. Notably, using MRC for measurement adds considerable load to the device, artificially inflating the reported latency.

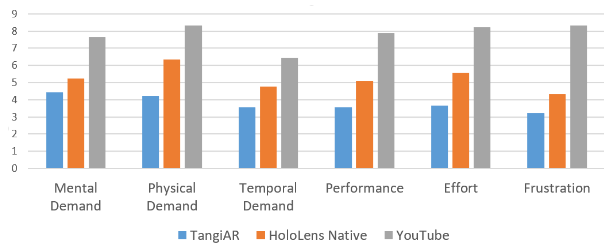
This latency falls within acceptable ranges for many AR interactions but approaches the threshold where users may notice lag in fast movements. Camera capture dominates the latency budget (120-140ms), highlighting that the tracking algorithm itself is sufficiently fast for real-time use. The remaining components—tracking computation (1-3ms) and network transmission (2-3ms)—are negligible, suggesting that hardware improvements rather than algorithmic optimization will drive future latency reductions.

8.4 Usability

The sub-par YouTube experience on HoloLens stems from its UI being adapted from tablet controls rather than optimized for immersive AR. This raises an important question: how should UI be



(a) Time used for task completion in seconds



(b) NASA-TLX

Figure 8: The results from the video player task

redesigned for HoloLens and similar HMDs? Many users noted the buttons felt too small—not an issue on tablets—due to a mismatch between system fingertip detection and user perception. Enlarging buttons could help but risks cluttering the interface. In contrast, *TangiAR* tangibles can occupy larger surfaces, offering more comfortable interaction space.

The HoloLens native player seems to acknowledge these challenges. By simplifying functionality, it allocates more space for buttons, and its use of sliding gestures helps reduce mis-triggers. Still, issues persist. Some users reported errors in precise progress control, where depth sensing caused inputs to shift upon finger release due to crossing gesture activation.

As for *TangiAR*, the results aligned with expectations. Tangible controls are known to be accurate, fast, and low-effort, and our findings reinforce this. Meanwhile, the study helped surface usability issues in our system, which we discuss in the Limitations section.

9 Limitations

9.1 Tracking Performance

While our system builds on a modern tracking algorithm, reliably detecting object pose in all conditions remains challenging. As observed in the user study, fast motion and significant occlusion can lead to tracking loss, which disrupts immersion. Fast movements may blur the image, making it harder for the system to recognize object shape. Hand occlusion presents a more fundamental challenge for tangible interaction, as our results show degraded performance beyond 50% occlusion levels. Despite these challenges, we find the current tracking performance sufficient to demonstrate the potential of markerless interaction.

9.2 Hardware Constraints

Our wide-field-of-view prototype proves too heavy for comfortable extended wear, while the standalone version constrains users to the limited field of view of the HoloLens camera. Despite mitigation strategies like pause/resume tracking and boundary warnings, users must keep objects within the camera's view, which can break immersion during natural interactions. This fundamental constraint affects the fluidity of tangible AR experiences.

9.3 Touch Detection Sensitivity

Touch interactions with tangible objects were highly praised by users but remain sensitive to tracking accuracy of both the object and hand joints. The small distance threshold required for reliable touch detection creates a trade-off between false positives and missed interactions, occasionally causing mismatches between visual and tactile feedback.

10 Future Work

Future tracking algorithms should focus on enhanced occlusion robustness and motion blur tolerance. Novel tracking algorithms using deep learning have shown increasing tolerance of occlusion. In this project, we refrained from using a tracking method based on neural networks, as these approaches typically require additional computation resources for training and deployment on mobile devices such as AR headsets. However, we do expect with next-generation AR headsets may have dedicated computer vision processors to support the integration of computationally costly tracking methods. Future integration of more high-end hardware on AR headsets may also improve motion blur tolerance. High-speed cameras combined with active IR illumination could mitigate image blur during fast movements. Our study also revealed limitation in touch detection sensitivity. Touch detection could be significantly improved by incorporating fingertip kinetics (distance, velocity, acceleration) rather than relying solely on proximity thresholds [18]. In terms of empirical findings, our evaluation is currently limited by a small sample size (N=10), which may reduce statistical power and generalizability of our findings. In the future, longitudinal studies in a larger and more diverse scale are needed to establish the broader applicability of markerless tangible AR. Future work can also explore collaborative multi-user scenarios [29] and domain-specific applications beyond media control and gaming.

11 Conclusion

Recent research has shown that tangible interaction is fast, accurate, and natural for users. However, building a convenient, marker-free tangible system remains technically challenging. In this paper, we address this gap by introducing *TangiAR*, a user-friendly, markerless tangible control system that runs in real time on an AR HMD. Through 2 applications and user studies, we validated its tangible usability and offer insights for AR UI design. Through technical evaluation, we demonstrate system performance under various conditions. Finally, we outline future directions to enhance the system and further explore the potential of tangibles in AR. We've open-sourced *TangiAR* at https://github.com/UBC-X-Lab/TangiAR.git to support future research.

References

[1] [n. d.]. OpenSCAD: The Programmers Solid 3D CAD Modeller. https://openscad.org/. Accessed: 2023-08-02.

[2] Ananta Narayanan Balaji, Clayton Kimber, David Li, Shengzhi Wu, Ruofei Du, and David Kim. 2023. RetroSphere: Self-Contained Passive 3D Controller Tracking for Augmented Reality. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 4 (Jan. 2023), 157:1–157:36. doi:10.1145/3569479

[3] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. 2006. SURF: Speeded Up Robust Features. In *Computer Vision – ECCV 2006*, Aleš Leonardis, Horst Bischof, and Axel Pinz (Eds.). Vol. 3951. Springer Berlin Heidelberg, Berlin, Heidelberg, 404–417. doi:10.1007/11744023_32 Series Title: Lecture Notes in Computer Science.

[4] Rodney A. Berry, Mark Billinghurst, and J. Kurumisawa. 2001. Augmented reality interface for electronic music performance. *proceedings of the 9th International Conference on Human-Computer Interaction (HCI International 2001)* (Jan. 2001). https://www.academia.edu/22297724/Augmented_reality_interface_for_electronic_music_performance

[5] Lonni Besançon, Paul Issartel, Mehdi Ammi, and Tobias Isenberg. 2017. Usability Comparison of Mouse, Touch and Tangible Inputs for 3D Data Manipulation. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 4727–4740. doi:10.1145/3025453.3025863 arXiv:1603.08735 [cs].

[6] M. Billinghurst, H. Kato, and I. Poupyrev. 2001. The MagicBook - moving seamlessly between reality and virtuality. *IEEE Computer Graphics and Applications* 21, 3 (May 2001), 6–8. doi:10.1109/38.920621 Conference Name: IEEE Computer Graphics and Applications.

[7] Mark Billinghurst, Hirokazu Kato, and Ivan Poupyrev. 2008. Tangible augmented reality. *ACM SIGGRAPH ASIA 2008 Courses* (Jan. 2008). doi:10.1145/1508044.1508051

[8] Mark Billinghurst, Ivan Poupyrev, Hirokazu Kato, and Richard May. 2000. Mixing Realities in Shared Space: An Augmented Reality Interface for Collaborative Computing., Vol. 3. 1641–1644. doi:10.1109/ICME.2000.871085

[9] Evren Bozgeyikli and Lal Lila Bozgeyikli. 2021. Evaluating Object Manipulation Interaction Techniques in Mixed Reality: Tangible User Interfaces and Gesture. In *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*. 778–787. doi:10.1109/VR50410.2021.00105 ISSN: 2642-5254.

[10] Fei Chen, Xiaodong Wang, Yunxiang Zhao, Shaohe Lv, and Xin Niu. 2022. Visual object tracking: A survey. *Computer Vision and Image Understanding* 222 (Sept. 2022), 103508. doi:10.1016/j.cviu.2022.103508

[11] Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. 2010. iCon: utilizing everyday objects as additional, auxiliary and instant tabletop controllers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '10)*. Association for Computing Machinery, New York, NY, USA, 1155–1164. doi:10.1145/1753326.1753499

[12] D. Comaniciu, V. Ramesh, and P. Meer. 2000. Real-time tracking of non-rigid objects using mean shift. In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, Vol. 2. 142–149 vol.2. doi:10.1109/CVPR.2000.854761 ISSN: 1063-6919.

[13] Christian Corsten, Ignacio Avellino, Max Möllers, and Jan Borchers. 2013. Instant user interfaces: repurposing everyday objects as input devices. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*. ACM, St. Andrews Scotland, United Kingdom, 71–80. doi:10.1145/2512349.2512799

[14] Adam Drogemuller, James Walsh, Ross T. Smith, Matt Adcock, and Bruce H Thomas. 2021. Turning everyday objects into passive tangible controllers. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3430524.3442460

[15] Adam Drogemuller, James Walsh, Ross T. Smith, Matt Adcock, and Bruce H Thomas. 2021. Turning everyday objects into passive tangible controllers. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '21)*. Association for Computing Machinery, New York, NY, USA, 1–4. doi:10.1145/3430524.3442460

[16] 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 *CHI Conference on Human Factors in Computing Systems Extended Abstracts*. ACM, New Orleans LA USA, 1–4. doi:10.1145/3491101.3519911

[17] Andreas Dünser, Julian Looser, Raphaël Grasset, Hartmut Seichter, and Mark Billinghurst. 2010. Evaluation of Tangible User Interfaces for Desktop AR. In *2010 International Symposium on Ubiquitous Virtual Reality*. 36–39. doi:10.1109/ISUVR.2010.19

[18] Neil Xu Fan and Robert Xiao. 2022. Reducing the Latency of Touch Tracking on Ad-hoc Surfaces. *Proceedings of the ACM on Human-Computer Interaction* 6, ISS (Nov. 2022), 577:489–577:499. doi:10.1145/3567730

[19] Martin Feick, Scott Bateman, Anthony Tang, André Miede, and Nicolai Marquardt. 2020. Tangi: Tangible Proxies For Embodied Object Exploration And Manipulation In Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 195–206. doi:10.1109/ISMAR50242.2020.00042 ISSN: 1554-7868.

[20] Thomas Feix, Javier Romero, Heinz-Bodo Schmiedmayer, Aaron M. Dollar, and Danica Kragic. 2016. The GRASP Taxonomy of Human Grasp Types. *IEEE Transactions on Human-Machine Systems* 46, 1 (Feb. 2016), 66–77. doi:10.1109/THMS.2015.2470657

[21] Markus Funk, Oliver Korn, and Albrecht Schmidt. 2014. An augmented workplace for enabling user-defined tangibles. In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. ACM, Toronto Ontario Canada, 1285–1290. doi:10.1145/2559206.2581142

[22] Christian Geiger, Leif Oppermann, and Christian Reimann. 2003. 3D-registered interaction-surfaces in augmented reality space. 5–13. doi:10.1109/ART.2003.1320417

[23] Mac Greenslade, Adrian Clark, and Stephan Lukosch. 2022. User-Defined Interaction Using Everyday Objects for Augmented Reality First Person Action Games. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. 842–843. doi:10.1109/VRW55335.2022.00272

[24] Robert Held, Ankit Gupta, Brian Curless, and Maneesh Agrawala. 2012. 3D puppetry: a kinect-based interface for 3D animation. In *Proceedings of the 25th annual ACM symposium on User interface software and technology (UIST '12)*. Association for Computing Machinery, New York, NY, USA, 423–434. doi:10.1145/2380116.2380170

[25] Steven J. Henderson and Steven Feiner. 2008. Opportunistic controls: leveraging natural affordances as tangible user interfaces for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology (VRST '08)*. Association for Computing Machinery, New York, NY, USA, 211–218. doi:10.1145/1450579.1450625

[26] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing Reality: Enabling Opportunistic Use of Everyday Objects as Tangible Proxies in Augmented Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, San Jose California USA, 1957–1967. doi:10.1145/2858036.2858134

[27] Yang Hua, Karteek Alahari, and Cordelia Schmid. 2015. Online Object Tracking With Proposal Selection. 3092–3100. https://openaccess.thecvf.com/content_iccv_2015/html/Hua_Online_Object_Tracking_ICCV_2015_paper.html

[28] Xincheng Huang, Keylonnie L. Miller, Alanson P. Sample, and Nikola Banovic. 2023. StructureSense: Inferring Constructive Assembly Structures from User Behaviors. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 6, 4, Article 204 (Jan. 2023), 25 pages. doi:10.1145/3570343

[29] Xincheng Huang and Robert Xiao. 2024. SurfShare: Lightweight Spatially Consistent Physical Surface and Virtual Replica Sharing with Head-mounted Mixed-Reality. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 7, 4, Article 162 (Jan. 2024), 24 pages. doi:10.1145/3631418

[30] Hiroshi Ishii and Brygg Ullmer. 1997. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems (CHI '97)*. Association for Computing Machinery, New York, NY, USA, 234–241. doi:10.1145/258549.258715

[31] Shahram Izadi, David Kim, Otmar Hilliges, David Molyneaux, Richard Newcombe, Pushmeet Kohli, Jamie Shotton, Steve Hodges, Dustin Freeman, Andrew Davison, and Andrew Fitzgibbon. 2011. KinectFusion: Real-time 3D Reconstruction and Interaction Using a Moving Depth Camera. In *UIST '11 Proceedings of the 24th annual ACM symposium on User interface software and technology*. ACM, 559–568. https://www.microsoft.com/en-us/research/publication/kinectfusion-real-time-3d-reconstruction-and-interaction-using-a-moving-depth-camera/

[32] Sergi Jordà, Günter Geiger, Marcos Alonso, and Martin Kaltenbrunner. 2007. The reacTable: exploring the synergy between live music performance and tabletop tangible interfaces. In *Proceedings of the 1st international conference on Tangible and embedded interaction (TEI '07)*. Association for Computing Machinery, New York, NY, USA, 139–146. doi:10.1145/1226969.1226998

[33] Michael Krainin, Peter Henry, Xiaofeng Ren, and Dieter Fox. 2011. Manipulator and object tracking for in-hand 3D object modeling. *The International Journal of Robotics Research* 30, 11 (Sept. 2011), 1311–1327. doi:10.1177/0278364911403178 Publisher: SAGE Publications Ltd STM.

[34] Gun Lee, Mark Billinghurst, and Gerard Kim. 2004. Occlusion based interaction methods for tangible augmented reality environments. 419–426. doi:10.1145/1044588.1044680

[35] Gun Lee, Claudia Nelles, Mark Billinghurst, and Gerard Kim. 2004. Immersive Authoring of Tangible Augmented Reality Applications. 172–181. doi:10.1109/ISMAR.2004.34

[36] Bo Li, Junjie Yan, Wei Wu, Zheng Zhu, and Xiaolin Hu. 2018. High Performance Visual Tracking With Siamese Region Proposal Network. 8971–8980. https://openaccess.thecvf.com/content_cvpr_2018/html/Li_High_Performance_Visual_CVPR_2018_paper.html

[37] Xi Li, Anthony Dick, Hanzi Wang, Chunhua Shen, and Anton van den Hengel. 2011. Graph mode-based contextual kernels for robust SVM tracking. In *2011 International Conference on Computer Vision*. 1156–1163. doi:10.1109/ICCV.2011.6126364 ISSN: 2380-7504.

[38] Rong-Hao Liang, Kai-Yin Cheng, Liwei Chan, Chuan-Xhyuan Peng, Mike Y. Chen, Rung-Huei Liang, De-Nian Yang, and Bing-Yu Chen. 2013. GaussBits:

magnetic tangible bits for portable and occlusion-free near-surface interactions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 1391–1400. doi:10.1145/2470654.2466185

[39] Julian Looser, Mark Billinghurst, and Andy Cockburn. 2004. Through the looking glass: the use of lenses as an interface tool for Augmented Reality interfaces. In *Proceedings of the 2nd international conference on Computer graphics and interactive techniques in Australasia and South East Asia (GRAPHITE '04)*. Association for Computing Machinery, New York, NY, USA, 204–211. doi:10.1145/988834.988870

[40] Manolis Lourakis and Xenophon Zabulis. 2013. Model-Based Pose Estimation for Rigid Objects. In *Computer Vision Systems (Lecture Notes in Computer Science)*, Mei Chen, Bastian Leibe, and Bernd Neumann (Eds.). Springer, Berlin, Heidelberg, 83–92. doi:10.1007/978-3-642-39402-7_9

[41] David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (Nov. 2004), 91–110. doi:10.1023/B:VISI.0000029664.99615.94

[42] C. McDonald and G. Roth. 2003. Replacing a Mouse with Hand Gesture in a Plane-Based Augmented Reality System. https://www.semanticscholar.org/paper/Replacing-a-Mouse-with-Hand-Gesture-in-a-Augmented-McDonald-Roth/83ce596c1b867b142a18879f1c5a406e1fbc02e1

[43] Naoki Numaguchi, Atsushi Nakazawa, Takaaki Shiratori, and Jessica K. Hodgins. 2011. A puppet interface for retrieval of motion capture data. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation (SCA '11)*. Association for Computing Machinery, New York, NY, USA, 157–166. doi:10.1145/2019406.2019427

[44] Ivan Poupyrev, D.S. Tan, Mark Billinghurst, Hirokazu Kato, Holger Regenbrecht, and Nobuji Tetsutani. 2002. Developing a generic augmented-reality interface. *Computer* 35 (April 2002), 44–50. doi:10.1109/2.989929

[45] Victor Prisacariu. 2009. PWP3D: Real-time segmentation and tracking of 3D objects. (Jan. 2009).

[46] Jan Riemann, Martin Schmitz, Alexander Hendrich, and Max Mühlhäuser. 2018. FlowPut: Environment-Aware Interactivity for Tangible 3D Objects. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 1 (March 2018), 31:1–31:23. doi:10.1145/3191763

[47] Julius Cosmo Romeo Rudolph, David Holman, Bruno De Araujo, Ricardo Jota, Daniel Wigdor, and Valkyrie Savage. 2022. Sensing Hand Interactions with Everyday Objects by Profiling Wrist Topography. In *Sixteenth International Conference on Tangible, Embedded, and Embodied Interaction (TEI '22)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3490149.3501320

[48] Kadek Ananta Satriadi, Jim Smiley, Barrett Ens, Maxime Cordeil, Tobias Czauderna, Benjamin Lee, Ying Yang, Tim Dwyer, and Bernhard Jenny. 2022. Tangible Globes for Data Visualisation in Augmented Reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)*. Association for Computing Machinery, New York, NY, USA, 1–16. doi:10.1145/3491102.3517715

[49] Manuel Stoiber, Martin Pfanne, Klaus H. Strobl, Rudolph Triebel, and Alin Albu-Schäffer. 2022. SRT3D: A Sparse Region-Based 3D Object Tracking Approach for the Real World. *International Journal of Computer Vision* 130, 4 (April 2022), 1008–1030. doi:10.1007/s11263-022-01579-8

[50] Manuel Stoiber, Martin Sundermeyer, and Rudolph Triebel. 2022. Iterative Corresponding Geometry: Fusing Region and Depth for Highly Efficient 3D Tracking of Textureless Objects. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, New Orleans, LA, USA, 6845–6855. doi:10.1109/CVPR52688.2022.00673

[51] Arda Ege Unlu and Robert Xiao. 2021. PAIR: Phone as an Augmented Immersive Reality Controller. In *Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21)*. Association for Computing Machinery, New York, NY, USA, 1–6. doi:10.1145/3489849.3489878

[52] L. Vacchetti, V. Lepetit, and P. Fua. 2004. Stable real-time 3D tracking using online and offline information. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 26, 10 (Oct. 2004), 1385–1391. doi:10.1109/TPAMI.2004.92 Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

[53] Virag Varga, Gergely Vakulya, Benjamin Buergisser, Nathan Riopelle, Fabio Zund, Robert W. Sumner, Thomas R. Gross, and Alanson Sample. 2021. Real-Time Capture of Holistic Tangible Interactions. In *Proceedings of the Fifteenth International Conference on Tangible, Embedded, and Embodied Interaction*. ACM, Salzburg Austria, 1–15. doi:10.1145/3430524.3440658

[54] James A. Walsh, Stewart von Itzstein, and Bruce H. Thomas. 2014. Ephemeral interaction using everyday objects. In *Proceedings of the Fifteenth Australasian User Interface Conference - Volume 150 (AUIC '14, Vol. 150)*. Australian Computer Society, Inc., AUS, 29–37.

[55] Qiang Wang, Zhu Teng, Junliang Xing, Jin Gao, Weiming Hu, and Stephen Maybank. 2018. Learning Attentions: Residual Attentional Siamese Network for High Performance Online Visual Tracking. 4854–4863. https://openaccess.thecvf.com/content_cvpr_2018/html/Wang_Learning_Attentions_Residual_CVPR_2018_paper.html

[56] Andrew D. Wilson. 2010. Using a depth camera as a touch sensor. In *ACM International Conference on Interactive Tabletops and Surfaces (ITS '10)*. Association for Computing Machinery, New York, NY, USA, 69–72. doi:10.1145/1936652.1936665

[57] Ziyi Xia, Xincheng Huang, Sidney S Fels, and Robert Xiao. 2025. HaloTouch: Using IR Multi-Path Interference to Support Touch Interactions with General Surfaces. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 548, 17 pages. doi:10.1145/3706598.3714179

[58] Robert Xiao, Chris Harrison, and Scott E. Hudson. 2013. WorldKit: rapid and easy creation of ad-hoc interactive applications on everyday surfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. Association for Computing Machinery, New York, NY, USA, 879–888. doi:10.1145/2470654.2466113

[59] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D. Wilson, and Hrvoje Benko. 2018. MRTouch: Adding Touch Input to Head-Mounted Mixed Reality. *IEEE Transactions on Visualization and Computer Graphics* 24, 4 (April 2018), 1653–1660. doi:10.1109/TVCG.2018.2794222 Conference Name: IEEE Transactions on Visualization and Computer Graphics.

[60] Tianyu Yang and Antoni B. Chan. 2017. Recurrent Filter Learning for Visual Tracking. 2010–2019. https://openaccess.thecvf.com/content_ICCV_2017_workshops/w28/html/Yang_Recurrent_Filter_Learning_ICCV_2017_paper.html

[61] Hui-Shyong Yeo, Ryosuke Minami, Kirill Rodriguez, George Shaker, and Aaron Quigley. 2018. Exploring Tangible Interactions with Radar Sensing. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 4 (Dec. 2018), 200:1–200:25. doi:10.1145/3287078

[62] Yunhua Zhang, Lijun Wang, Jinqing Qi, Dong Wang, Mengyang Feng, and Huchuan Lu. 2018. Structured Siamese Network for Real-Time Visual Tracking. 351–366. https://openaccess.thecvf.com/content_ECCV_2018/html/Yunhua_Zhang_Structured_Siamese_Network_ECCV_2018_paper.html