Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality

Rahul Jain*
jain348@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Jingyu Shi*
shi537@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Runlin Duan
duan92@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Zhengzhe Zhu
zhu714@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Xun Qian
qian85@purdue.edu
Purdue University
West Lafayette, Indiana, USA

Karthik Ramani
ramani@purdue.edu
Purdue University
West Lafayette, Indiana, USA

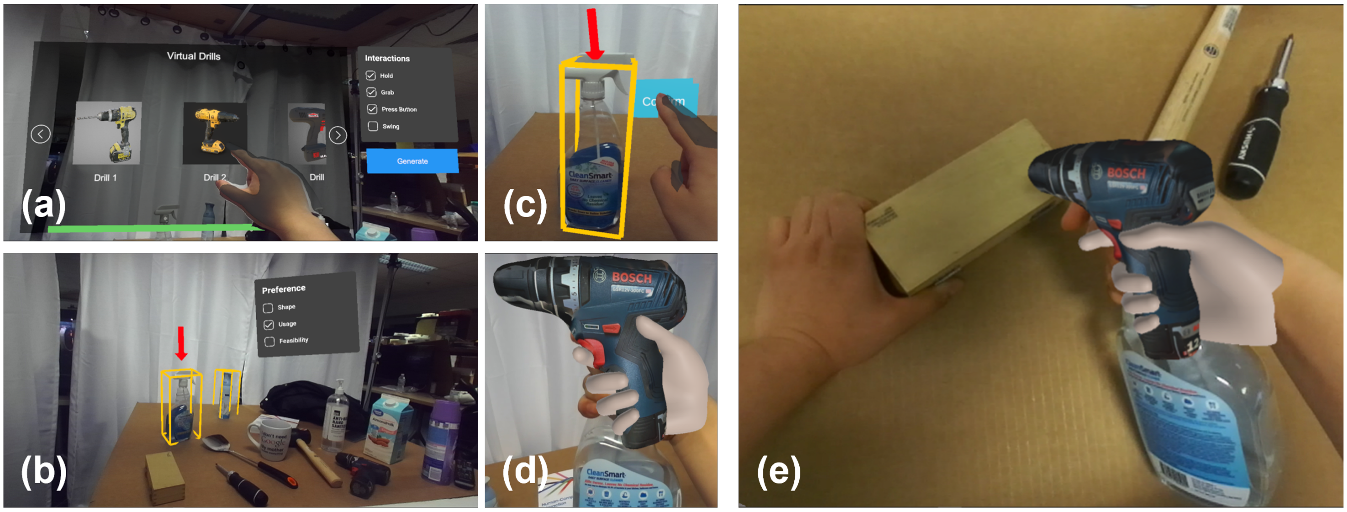


Figure 1: An overview of Ubi-TOUCH system workflow. The user of Ubi-TOUCH needs to interact with a virtual object in an AR application as input, in this figure, e.g., practicing using a drill. There is no drill handy in the user's vicinity. To use Ubi-TOUCH to find a tangible proxy, they first select the target virtual object (a). Then the system scans the environment and recommends the physical objects within as the proxies to the user (b). The user selects a physical object (c) and Ubi-TOUCH aligns the virtual and the physical objects, maps the physical hand-object interaction to a consistent virtual hand-object interaction, and blends the interaction into reality (d). Finally, Ubi-TOUCH enables the user to interact with the virtual object with haptic feedback and consistent visualization of the hand-object interaction.

ABSTRACT

Utilizing everyday objects as tangible proxies for Augmented Reality (AR) provides users with haptic feedback while interacting with virtual objects. Yet, existing methods focus on the attributes of the objects, constraining the possible proxies and yielding inconsistency in user experience. Therefore, we propose Ubi-TOUCH, an AR system that assists users in seeking a wider range of tangible proxies for AR applications based on the hand-object interaction (HOI) they desire. Given the target interaction with a virtual object, the system scans the users' vicinity and recommends object proxies with similar interactions. Upon user selection, the system simultaneously tracks and maps users' physical HOI to the virtual HOI, adaptively optimizing object 6 DoF and the hand gesture to provide consistency between the interactions. We showcase promising use cases of Ubi-TOUCH, such as remote tutorials, AR gaming, and Smart Home control. Finally, we evaluate the performance and usability of Ubi-TOUCH with a user study.

CCS CONCEPTS

• **Human-centered computing → Visualization systems and tools.**

*Both authors contributed equally to this research.

UIST '23, October 29-November 1, 2023, San Francisco, CA, USA
© 2023 Copyright held by the owner/author(s).
ACM ISBN 979-8-4007-0132-0/23/10.
https://doi.org/10.1145/3586183.3606793

KEYWORDS

Augmented Reality, Opportunistic Tangible Proxy, Tangible User Interface

ACM Reference Format:
Rahul Jain, Jingyu Shi, Runlin Duan, Zhengzhe Zhu, Xun Qian, and Karthik Ramani. 2023. Ubi-TOUCH: Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in Augmented Reality. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29-November 1, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3586183.3606793

1 INTRODUCTION

In Augmented Reality (AR), interacting with virtual components lacks haptic feedback [10, 47, 89]. To address this issue, a handful of approaches have been studied to enable tangible AR applications, such as designing wearable hardware [19, 64, 77, 99], retargeting to self-haptics [30, 74], and programming tangible input devices [2, 20, 26, 32, 33, 87]. Recent research on retargeting everyday objects as tangible proxies shows promising results in natural, intuitive, and inclusive interactions with the virtual components [7, 40, 43, 82]. By opportunistically repurposing and leveraging the existing physical objects in the users' environment as input devices, users are freed from hardware constraints and obtain realistic haptic feedback within the AR experience.

Precise mappings are crucial to the correspondence between everyday physical objects and their intended virtual counterparts to produce interactions that are both physically and mentally aligned [15, 61] for users. Such mappings must satisfy both the geometrical and semantic constraints [42, 62, 98] of the components. For example, a cell phone would not be preferred as a proxy for a basketball, since neither do they share the same geometric attributes, nor are they used for similar purposes. Thus, formulating reliable mapping criteria is a significant challenge in the investigation of opportunistic tangible proxies.

Prior research put considerable effort into addressing this challenge. Annexing Reality [42] enables the users to define preference in geometric shape and matches the given virtual object with physical objects in the vicinity that are most similar in the preferred geometric shape. Inspired by this work, following-up research like [28, 44, 98] seeks opportunistic proxy objects by matching the physical attributes of the objects in the interaction. While successfully providing the best-available haptic sensation for virtual objects, such methods put heavy constraints on the physical attributes of the objects and thus restrict the possible range of opportunistic proxies. For instance, a proxy for a virtual basketball would always be a sphere regardless of the affordance of the basketball. Affordance should also be one of the criteria while matching between objects [46]. To this end, [29, 31, 41, 62] take the affordances of the objects into consideration, matching proxies based on whether they can be used in the same manner. However, the inconsistency in the shapes of the objects results in the Breaks In Presence (BIP) [21, 53, 81, 85] in the user experience and consequently defects in the efficiency of the interaction [52]. BIP happens when the proxies have different geometry from the virtual counterparts, resulting in the users interacting with the objects while seeing their physical hands inconsistently penetrating, isolated from, or blocked by the virtual overlays. Moreover, not all affordances are needed for a particular interaction, i.e., the constraints on the affordance of the objects should be decided by the intended interactions [35, 69, 94]. Recent work by [69] proposes to capture real-world interactions and prototype user-defined AR applications, allowing flexible and general-purpose AR prototypes. Enlightened by these works, we aim to address this dilemma between the constraints on object selection and the inconsistency of the user experience.

To this end, we propose Ubi-TOUCH, an AR system for empowering Ubiquitous Tangible Object Utilization through Consistent Hand-object interaction in AR. Given target interactions with a virtual object in AR, Ubi-TOUCH recommends the best-available proxies in users' vicinity for target interactions, and, per user selection, maps the real-world Hand-Object Interactions (HOIs) to the virtual HOIs, and provides consistent visualization of the interaction to the users. Instead of merely focusing on the object attributes such as shape, size, and affordance, Ubi-TOUCH considers attributes of HOIs, motivated by the fact that the HOIs are essentially the bridge between end-users and the AR applications. Ubi-TOUCH keeps the physical and mental consistency in user experience by opting for objects by interaction constraints. We utilize a comprehensive taxonomy of HOIs to propose possible mappings to the users to also enable greater flexibility and generalizability in seeking opportunistic proxies.

We develop an integrated vision-based workflow to firstly, establish a knowledge base of HOI, containing object-wise interaction attributes such as contact points, affordance, and hand poses secondly, scan the end-user environment, detect objects, and recommend the best proxies in the environment based on the similarity between interactions with the target object and those with the possible objects (Figure 1 (b)) thirdly upon user selection, track the hand and object and simultaneously map the physical interaction into the virtual space (Figure 1 (d)) as possible inputs to any AR application (Figure 1 (e))

We list our contributions in the following:

- A comprehensive vision-based workflow that assists AR users in finding everyday objects as opportunistic tangible proxies based on hand-object interaction constraints,
- A contact-point-based optimization technique to render hand-object interaction with consistency among different objects,
- An AR interface that enables hand-object interaction with tangible proxies for virtual objects, incorporating real-time virtual interaction blending.

2 RELATED WORK

2.1 Tangible AR

Tangible augment reality (TAR) [11] seeks the seamless interactions between the virtual and the physical world by combining the display possibility of AR and the haptic feedback of the physical objects. Early work on the TAR augments the virtual objects and information on the AR marker cards [38, 76]. Illuminating clay [75] introduces an alternative modeling material as a physical proxy of the virtual terrain for landscape analysis. Rubik's cube [9] proposes AR Rubik's cues as a controller and game board for AR gameplay. Other input modalities are adopted to the TAR systems, like the paddle-like proxies for virtual object interactions [51] [66] and gesture recognition on AR markers to create tangible AR experiences [66] [84]. Lee et al. [60] suggest an occlusion-based interaction in which the visual occlusion of the physical markers triggers the two-dimensional interactions. More physical objects from the domain environment are adopted to the TAR. Oda et al. [71] develop a car racing system using a pre-manufactured passive tangible controller as input. Knoerlein et al. [54] present a collaborative AR ping-pong system that is supported by virtual bats colocated with haptic devices.

The in-efficiency and low resolution of the early AR devices limit the TAR system input on markers and pre-defined physical proxies. Moreover, the intimate relationship between the digital models and the physical objects reveals the advantage of the intuitive consistency of everyday objects to provide haptic feedback as tangible proxies.

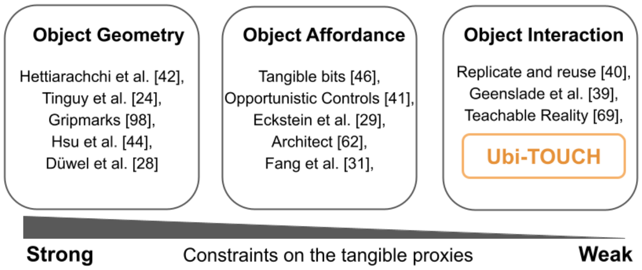


Figure 2: Three criteria for seeking opportunistic objects as tangible proxies for AR. *Object Geomerty*: The geometric features (e.g. shape, size, surface) are strictly similar to the target object. *Object Affordance*: The use of the object or the action possibility is similar between the object and the proxy. *Object Interaction*: Both possible actions and the experience of the users (e.g. gesture and movement of the body) are aligned between the target object and the proxy.

2.2 Opportunistic Objects as Tangible Proxies

Tangible proxies are physical representations of virtual objects that facilitate haptic feedback for humans while maintaining natural and intuitive manipulation experiences. Opportunistic controls [41] suggest using the existing opportunistic object in the domain environment as tangible proxies for the user inputs of AR applications. By transforming opportunistic objects [42] into tangible proxies, users can experience greater flexibility as they are no longer confined to the limitations of physical objects with pre-defined digital functions.

Prior efforts to address this challenge can be classified into three primary categories: 1) tangible proxies that base on the geometry primitives of objects, 2) those that consider object affordances, and 3) those that emphasize object interactions.

Object geometry primitives: Finding the tangible proxy based on the geometry primitives is inspired by the work of studying the object similarity on the user's interaction experience. Though the slight mismatch in the geometry primitives is acceptable, the lower disparity improves the manipulation quality and users' believability [12, 24]. Another research illustrates that illusionary haptic sensations happen due to the dominance of visual stimuli over kinesthetic stimuli [92]. Based on these elicitation studies, Hettiarachchi et al. [42] propose that the geometry primitives can represent the opportunistic object for tangible proxy. Following their work, Tinguy et al. [25] propose a new approach to take haptic sensations into consideration to provide a better match for the tangible proxies. The Gripmarks [98] develop a system that enables users to adopt opportunistic objects they already hold as input surfaces. Hsu et al. [44] present a prototype system that transforms the physical object into virtual models. Another way of using the geometry primitives includes using reconfigurable tangible proxies. Düwel et al. [28] propose to utilize interaction information to aid in geometric primitive matching. VirtualBricks [4] proposes a LEGO based toolkit enabling controllers for VR to simulate a variety of physical-manipulation. Ruofei et al. [27] study the possibility to transform everyday objects in to tangible interfaces based on their 6 DoF. Florian et al. [22] study the same case as in VR.

Object affordances: The object affordances of the physical object can be used to match the tangible proxy of the virtual object. Tangible bits [46] suggest using the natural object's physical affordance to bridge the physical and virtual interactions. Opportunistic Controls [41] first leverage the affordance of the domain object to provide tangible feedback for augment reality applications. Eckstein et al. [29] implement a prototype that substitutes the real world with a virtual environment by regulating the mismatch of the affordance between virtual and physical objects. ARchitect [62] considers the interactive affordances of the physical object and utilizes the knowledge to generate a VR experience. In their work, opportunistic objects, such as chairs or umbrellas, are suggested by the system as the tangible proxy for the virtual scene based on corresponding affordances, such as sitting on or grabbing. Recently, Fang et al. [31] state that similar affordance of the objects could enhance the tactile feedback for tangible interaction. Consequently, they introduce a method of repurposing opportunistic objects in the home to provide tangible experiences considering the object's shape and affordance. Gripmarks [98] also leverage the object's inherent physical affordance for their input surface design.

Object interactions: Recent work investigates the interaction design of tangible proxies to enhance tangible AR experiences. Replicate and Reuse [40] overlaid digital information on tangible physical objects and investigated the interaction designs of three physical objects with different tangibility levels. Greenslade et al. [39] present a study on utilizing everyday objects as tangible proxies for user-defined interactions in AR games. Teachable Reality [69] introduces an augmented reality (AR) prototyping tool that enables the creation of interactive tangible AR applications utilizing opportunistic objects. It lowers the barrier for AR prototyping by automating the identification of user-defined tangible proxies and facilitating gestural interactions through a computer vision model.

Tangible proxies driven by interactions offer advantages over geometry primitives and affordance proxies, as they enable more flexible and general-purpose AR prototypes. However, current work is facing the problem of low efficiency and over-constraints in finding the optimal tangible proxies due to the absence of deep insights into the interactions between humans and opportunistic objects.

2.3 Hand-Object Interaction in AR

Hand-object interaction (HOI), as the primary way humans interact with surrounding objects, provides intuitive, accessible, and authentic passive haptic feedback essential for creating tangible AR experiences.

Many works have investigated HOI in AR applications. Pei et al. [74] propose a new technique that transforms the users' hands into tangible proxies by imitating the virtual object themselves. They designed the hand posture of the target object by considering the shape similarity, comfort, and social acceptance. Virtual Grasp [95] allows users to retrieve virtual objects by performing the gestures they normally use for interacting with the physics object. Kosch et al. [55] create a tangible digital interface by incorporating everyday objects and their hand-object interactions. GesturAR [90] proposes an end-to-end AR authoring tool that incorporates customized hand interactions to enable users to create in-situ tangible AR applications through embodied demonstration and visual programming. CapturAR [91] enables users to author context-aware AR applications, employing hand-object interactions (HOI) to activate personalized functions. Fang et al. [30] propose an alternative approach that utilizes the users' bodies to generate haptic feedback for immersive experiences. AdaptutAR [45] offers learners an adaptive augmented reality learning experience tailored to their individual progress by tracking the interactions between users and machine interfaces. ARnnotate [78] enables users to create custom data for vision-based 3D hand-object interaction pose estimation using AR devices capable of hand tracking. Recent work, Ubi-edge [34] leverages hand interaction with object edges to create opportunistic tangible user interfaces in AR.

Drawing upon prior research on HOI in AR applications, Ubi-TOUCH specifically addresses the challenge of effectively identifying potential tangible proxies from opportunistic objects by considering the essential attributes of HOI, such as contact points, affordance, and hand pose. Leveraging a comprehensive taxonomy of HOI, our system offers enhanced flexibility and generalizability in locating opportunistic proxies, thereby facilitating intuitive and proficient tangible AR interactions

3 DESIGN RATIONALE

3.1 Hand-Object Interaction

HOI is an essential aspect of human activity, and we use our hands to manipulate and interact with objects in our environment on a daily basis. HOI can involve many actions, such as picking up objects, using tools, performing deictic gestures, etc. HOI has also become increasingly vital in the digital realm, with the development of AR and other immersive technologies [23]. The range of HOI is enlarged when we blend the virtual and physical worlds. HOIs are composed of hand gestures and their action on the objects as well as the contact points on both the hands and the objects. Consider two different interactions of a hand and a fork: 1) using a fork to eat, and 2) handing it off. The grasping gesture, contact points on the hand and the object, as well as the object 6 DoF, are different between these two interactions. As shown in Figure 3, people grasp a fork by the handle to use it but grasp the fork by the fork side to hand it off. HOIs are described by which affordances of the objects are handled by the hands and how the hands(gesture) handle the objects. To this end, we classify HOIs into two dimensions as shown in Figure 4. The first dimension is the movement of the HOI:

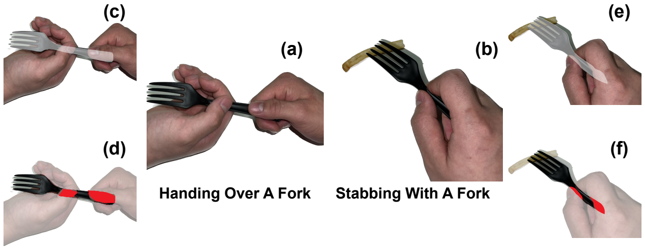


Figure 3: A decomposition of two HOIs with a fork. HOIs (a,b) are composed of the involved object (d,f), and the hand (c,e). The affordance of the object decides how the object is used, resulting in different contact points as shown in red. The gesture involved in the HOI also plays a role in determining the contact points as well as how the affordance of the objects is realized. A different gesture or a different object will make the HOI different.

- ***Static***: Interactions where the hand and object remain in a fixed position. It depicts scenarios where the location and orientation of the object remain unchanged while an intended interaction happens, such as pressing a button on a remote with no movement of the remote.
- ***Dynamic***: Interactions where the hand and object are in motion. This requires the hand to manipulate the object in a way that changes its position or orientation, often involving grasping, lifting, or cutting actions that change the object's 6 DoF.

The second dimension is the contact time of the HOI:

- ***Transient***: Interactions where the contact between hand and object is for a very short time. The contact between hand and object is very brief and often contains rapid movements.
- ***Continuous***: Interactions where the hand remains in contact with an object for a longer period.

Currently, we consider both articulate and non-articulate objects as rigid entities. Each object has only one unique center. This center determines whether the interaction is dynamic or static regardless of the hand movement, in order to avoid the complexity induced by the articulation of the object.

3.2 Hand-object Interaction Mapping

Mapping refers to establishing a correspondence between similar modalities, which can be objects, gestures, and interactions. To use physical objects as the proxies for interacting with virtual objects, mappings are needed to keep the consistency between both physical and virtual interaction. We categorize HOI interactions into four categories mentioned in section 3.1. The categorization empowers the mapping between the HOI interactions by thresholding the search space for mapping. In other words, given a user-selected interaction, we only consider possible interactions of the same category for mapping. For example, a dynamic continuous interaction will only be mapped into another dynamic continuous interaction.

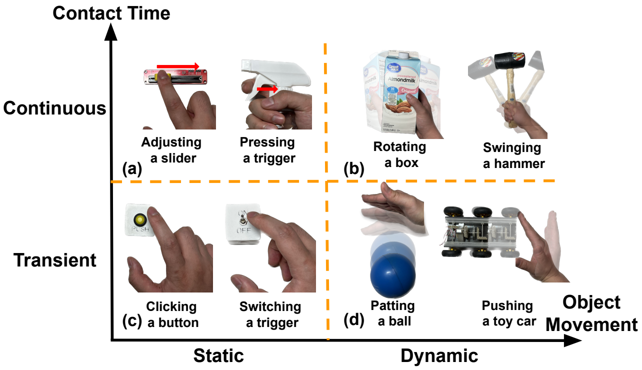


Figure 4: A taxonomy of HOI. Being either *Static* and *Dynamic* depict the movement of the object changes or not. Based on the length of the contact time, an HOI can be further described as *Continuous* (long and constant contact time) and *Transient* (short and discrete contact time).

We take into consideration the essential components of an HOI, namely the object, the hand gesture, and the contact points on both.

3.2.1 Mapping of object. As the criterion for mapping the objects, we consider both object geometry and object affordance.

Geometry Mapping Object manipulations are more efficient when physical and virtual objects are alike in shape and size [65, 68]. Geometric attributes of the objects such as shape, curves, size, curvature, and surface normals are used to map virtual objects to physical proxies to provide proximate haptic feedback to the users. We aim to utilize geometric features as one of the criteria to map physical objects and virtual objects. The similarity between the geometric features of the objects depicts naturally how similar two objects look and is able to enhance the immersiveness of the AR blending of the virtual object. Therefore, the more geometrically similar the objects are, the more plausible we consider this mapping.

Affordance Mapping We follow the definition of the terminology *affordance* as in [36], both actual and perceived possibilities of an object in relation to the action capabilities of an actor. In other words, the affordance of an object is what the user can do with it, whether intended or not. For example, in Figure 3 the fork can be *held* from the handle while the *stabbing* action is performed with the points of the fork. While opting for physical proxies for virtual objects, the similarity in object affordance often suggests a more natural substitute due to a similar spectrum of possibilities of actions [37].

In our design process, we utilize affordance as one of the criteria to map the objects, as the similarity in object affordance is crucial for creating a believable experience. For instance, a saw can be a better proxy for a virtual knife rather than a ruler. Even though a ruler shares similar geometry with a knife, it cannot cover the function of cutting like a knife, especially when the user wants to cut something with a knife. In terms of finding the proxy for a virtual object, the similarity in affordance is more important when the user is to perform an intended interaction with the virtual object. Overall, the concept of affordance is a critical attribute of both virtual and physical objects and is a key characteristic for mappings between HOIs to create a more realistic and immersive experience.

3.2.2 Gesture Mapping. Hand gestures provide a "hint" for the type of hand-object interactions to be performed [50]. The poking interaction of the fork (Figure 3) has a wrapping gesture of the hand which indicates the hand is holding something. Often, hand gestures vary with objects and the type of interactions. For instance, the hand gesture of *grabbing* a bottle is different from that of *grabbing* a cell phone, even though the interactions are both *grabbing*. The hand gesture of *grabbing* a bottle is different from that of *opening* the bottle, given the same objects interacted with. To this end, we include hand gestures also as one criterion to map one interaction to another. Intuitively, gestures should be mapped in interactions with similar poses.

3.2.3 Contact mapping. Contacts refer to the points on the objects and hands at which they touch each other during the interactions. Contact points are influenced by hand-object interaction. For example, contact points on the bottle cap and the base of the bottle signify two different interactions (i.e., opening the bottle and holding the bottle). Contact points on the object indicate the possible interaction performed with the object as well as the gestures. Hence, to map interactions, contact points should also be mapped from one object to another.

4 UBI-TOUCH SYSTEM

In this section, we walk through the system in more detail with an example. Then discuss the implementation of various modules of Ubi-TOUCH. Finally, we present our interface.

4.1 System Overview

The workflow of Ubi-TOUCH is composed of four steps as listed in the following:

Interaction Selection The user selects the interaction they want to perform with any virtual object. This piece of information is later used as the criterion for mapping the interactions, as described in Section 3.2.

Object Registration The system scans the physical environment, detects physical objects, and then registers the scanned objects. The categorization and geometric shape of the registered objects are taken into consideration when mapping the geometry and affordance of the HOI.

Proxy Recommendation The registered objects are evaluated based on the interaction knowledge and the target interaction given. Following our discussion in Section 3.2, we design our recommendation algorithms based on object geometry, object affordance, and hand gesture of the interaction with the proxies. A score of each possible proxy is computed and made visible to the user suggesting how similar the interaction with the proxy is to the intended interaction.

Interaction Mapping The user selects an object from the environment and interacts with the object. The system captures the physical interaction between the user and the object and then maps the interaction to the virtual counterpart.

To better understand the system, we elaborate with an example where the user wants to perform drilling interaction with a virtual drill. Specifically, the user wants to manipulate the drill to the desired location and press the trigger for drilling as shown in Figure 1. The system guides the user through the process of selecting a proxy object to map the hand-object interactions to the virtual drill. Firstly, the user selects the interactions they want to perform with the virtual drill, in this case, manipulating and pressing the trigger. Then the user scans the environment to register objects in the vicinity (Figure 1 (b)). Once the objects in the environment are registered, the system recommends objects to the user that can be used as a proxy for interacting with the virtual drill (Figure 1 (b)). In this case, the system recommends objects (a sprayer and a bottle) among all the objects in the environment, based on our proposed mapping criteria elaborated in the following subsections. Then the user selects the sprayer as the best proxy (Figure 1 (c)) and the system aligns the virtual drill to the sprayer based on contact point information. Eventually, the user interacts with the sprayer. The hand-object interactions with the sprayer are mapped to those with the virtual drill. The user can then learn to fix a glass box using the sprinkler as a proxy for the hand drill (Figure 1 (d)).

4.2 Interaction Selection

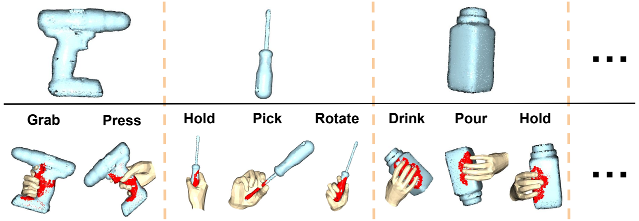


Figure 5: Some examples of the database of HOI. The first row is the 3D models of the objects. For each object, we collect their affordances, i.e. action possibilities, and compute their contact points with the possible hand gestures as shown in the second row for each interaction.

This module shows users virtual objects and the interactions they can perform with them based on a database of hand-object interactions. The database is created with various hand-object interaction data points following the taxonomy of HOIs discussed in 3.1. Each interaction in the database represents a single data point, consisting of information such as the interaction type, the 3D model of the object, the object's affordance (i.e., its intended uses), the hand pose required to perform the interaction, and the contact heat maps between the hand and the object as shown in Figure 5. The database is constructed from HOIs that commonly occur in daily life, and the sources for collecting the interactions include computer vision datasets such as ContactPose [13], GRAB [88], OakInk [96]and H20 [57]. Some examples of the interactions in the database are shown in Figure 5, illustrating different hand-object interactions for different types of objects.

To add more data points to the interactions database, a 5-camera hardware setup was constructed, as shown in Figure 6. We follow the approach in H20 [57] to process the images captured by the setup and obtain the hand poses, object poses, contact points, and type of interactions. Utilizing this setup and data collection approach we can capture interactions with new objects or objects that were not previously included in the database and thus further expand the interactions database and generalize the use cases.

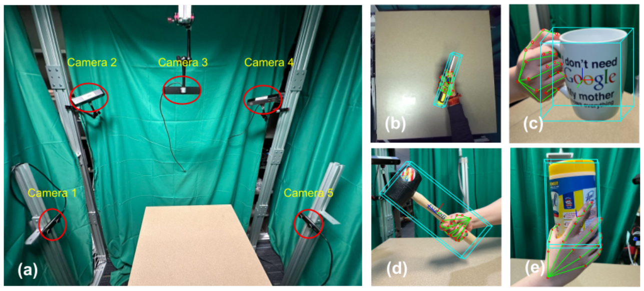


Figure 6: An illustration of our setup for collecting the HOI database. We utilize a 5-camera system to capture RGB-D images of hands interacting with various objects (a). We calibrate and leverage the multiple views to annotate the hand joints, object 6-DoF, and object bounding-boxes (b,c,d,e)

4.3 Object Registration

To register objects in the scenario, Ubi-TOUCH first detects them during scanning using a well-known RGB-based object detection method [79]. We obtain bounding boxes around the object and extract the object point cloud by projecting the bounding box to 3D and filtering the background points with the distance. Extracted object point cloud is used as an input for instance-level retrieval of the corresponding object model from the database using a deep learning-based 3D retrieval algorithm PointNet [18]. Note that we narrow the search range down to one category by object classification to reduce the retrieval time.

4.4 Proxy Recommendation

After registering all detected objects in the scenario and retrieving them from the database, Ubi-TOUCH computes the similarity between the target interaction selected by the user and the possible interactions with the registered objects. Following discussion in 3.2, we formulate the similarity among interactions by *Object Geometry*, *Object Affordance*, and *Hand Gesture*.

4.4.1 Object Geometry. We consider shape, curves, size, curvature, and surface normals for both the virtual object and physical objects registered during scanning. PointNet [18] is utilized in Ubi-TOUCH to compute the global geometric features such as coarse shape features given the point cloud of an object. We then compute the geometric features of the registered objects as well as that of the user-selected virtual object. Given two sets of geometric features of two objects respectively, we compute the Geometric similarity score by line 2 in Algorithm 1, where $O.Geo$ and $O_v.Geo$ are the extracted geometric features of the virtual object and those of the physical object respectively. After calculating the cosine similarities between objects in the database, objects are ranked based on their

Algorithm 1 Similarity Score Calculation

```
1:  for Each registered object $O$ do
2:      $O.Score_{geometry} \leftarrow cos(O_v.Geo, O.Geo)$
3:                                              ▷ Geometry Similarity
4:      $O.Score_{affordance} \leftarrow intersection(O_v.aff, O.aff)$
5:                                              ▷ Affordance Similarity
6:      for Each $O_v.I_v$ do
7:          for Each $O.I$ do
8:              $O.I.gesture_{old} \leftarrow O.I.gesture$
9:              $O.I.gesture \leftarrow$ Equation 6
10:                                             ▷ Optimize hand gesture
11:             $O_v.I_v.Score_{gesture} \leftarrow$
12:                 $cos(O.I.gesture, O.I.gesture_{old})$
13:                                             ▷ Gesture Similarity per interaction
14:         end for
15:     end for
16:     $O.Score_{gesture} \leftarrow average(O_v.I_v.Score_{gesture})$
17:                                             ▷ Gesture Similarity per object
18: end for
```

similarity scores in descending order. The objects with higher cosine similarity scores are more similar to the query object. We then consider the top-3 most similar objects from the ranked list as your final recommendation based on geometric features to the user.

4.4.2 Object Affordance. With the retrieved object information from the database, Ubi-TOUCH also evaluates the similarity between the affordance of both the target interaction and the potential interactions with the registered objects. Specifically, a list of interactions is created when the user selects the intended interactions with the virtual objects. For each registered object, we obtain another list of interactions from the database. We then compute the object affordance similarity between each registered object by the intersection of the two lists as shown in line 3 in Algorithm 1, where $O_v.aff$ and $O.aff$ are the list of interactions of the virtual object and that of the physical object respectively

4.4.3 Hand Gesture. As a crucial component of HOI, the gesture of the hand is key to evaluating the similarity between two interactions. For each user-selected interaction with the virtual object, we first retrieve the hand gesture and the contact heat map of this interaction. We transfer the contact heat map to the registered object to obtain the corresponding contact heat map of this interaction on the registered object (Figure 7). As shown from line 6 to line 10 in Algorithm 1, for each user-selected virtual object (denoted as O_v) and its possible virtual interaction (denoted as $O_v.I_v$), we paired them with each possible interaction with the registered object (denoted as $O.I$). We then optimize the hand gesture by Equation 6, (denoted as $O.I.gesture$). We further elaborate on the loss terms in 4.5. This optimization adapts the hand gesture to the target interaction with the registered object. We then compute the similarity score between the original hand gesture and the optimized hand gesture by calculating the cosine of them in lines 11 and 12 of Algorithm 1. We finally take the average across all gesture similarity scores and assign the score to the registered object O in line 14.

To this end, for all registered objects, we compute three different similarity scores for a target interaction. We suggested the best fit tangible proxy based on the similarity scores and the user's preference.

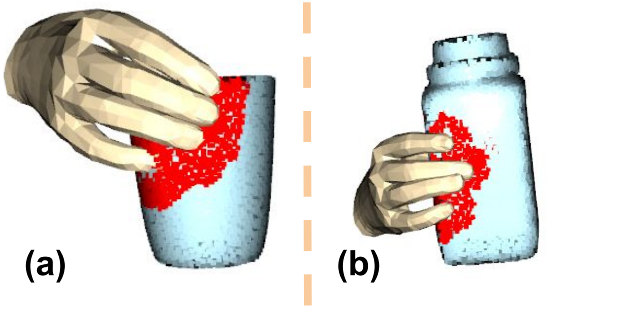


Figure 7: A visualization of contact heat maps transferring from one object to another. (a) shows the interaction of a hand holding a cup, and (b) shows the interaction of a hand holding a bottle. Different objects in HOI yield different contact points and different gestures. Our method aims to find the mapping between the two sets of contact points.

4.5 Tracking and Mapping

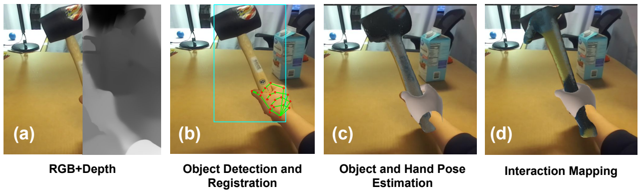

(a)
RGB+Depth
(b)
Object Detection and Registration
(c)
Object and Hand Pose Estimation
(d)
Interaction Mapping

Figure 8: An overview of our Tracking and Mapping pipeline. (a) We take RGB-D images as input. (b) We first detect and classify the objects from the input. After retrieval, registration, and user selection of the object as the proxy, we start tracking by estimating the physical world interaction (c, i.e. the object 6 DoF and mesh plus the hand pose and mesh). We keep optimizing the estimation from the physical world and map this interaction to the virtual world with the target object(d)

To accurately map the physical HOIs to the virtual counterpart, the movement of the hand, the object, and the points of contact between them should be tracked over time. This is done by detecting and tracking the position and orientation of the hand and the object in each frame of a video or sequence of images.

4.5.1 Hand Tracking. To detect and track the hand pose, we use a deep learning algorithm[80]. The biggest advantage provided by the algorithm is detecting hands in complex scenarios such as cluttered backgrounds, different lighting conditions, motion blur, and occlusion. we use a pre-trained algorithm on 100 DOH Dataset [83]. We perform hand tracking frame to get hand pose.

Algorithm 2 Tracking and Mapping

```
while Input Frame $f$ is valid do
    $V_f \leftarrow FrankMocap(f)$                                   ▷ Hand Mesh
    $R_f, T_f \leftarrow MegaPose(f)$                                ▷ Object 6 DoF
    $\hat{R}_f, \hat{T}_f, \hat{V}_f \leftarrow$ Equation 4
    $R_{v,f}, T_{v,f} \leftarrow R_f, T_f$               ▷ Mapping to Virtual Interaction
    $\hat{R_{v,f}}, \hat{T_{v,f}}, \hat{V_f} \leftarrow$ Equation 6      ▷ Optimize the Virtual Hand
    $\hat{R}_v, \hat{T}_v, \hat{V} \leftarrow Kalman\_Filter(\hat{R_{v,f}}, \hat{T_{v,f}}, \hat{V_f})$
                                                          ▷ Update the Kalman Filter
end while
```

4.5.2 Object Tracking. In order to track the object we use a generalized Deep-Learning-based algorithm MegaPose [59] to track the 6 DoF of an object in a video or sequence of images. MegaPose utilizes geometric and visual features from the input data to improve the accuracy of the 6 DOF predictions. After obtaining the initial results from MegaPose, we further refine the object pose using ICP [5]. Performing this step frame by frame ensures that the object pose is accurately tracked over time, and can be especially important when analyzing complex interactions between the object and other elements in the scene.

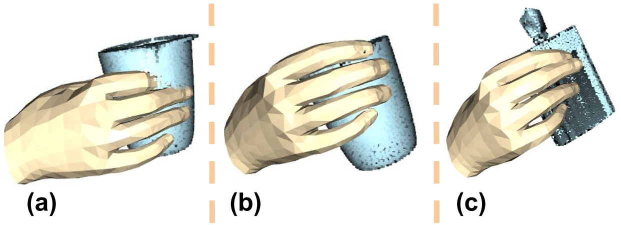


Figure 9: An illustration of our Optimization target. Given the estimated 6 Dof, 3D model of the physical object, and 3D hand mesh in the physical world, the original estimation results are shown in (a). The hand mesh is penetrating the object mesh which is not realistic. We optimize the joint pose to (b), correcting the hand pose and the distance between the object and the hand. We then map the interaction into the virtual world and render the same interaction but with a different object (c). We similarly optimize the virtual interaction while additionally adding constraints regarding the contact points, in order to get consistent virtual interactions.

4.5.3 Joint Hand Object Pose Optimization. Often separate tracking of both hands and objects results in implausible 3D reconstructions such as the object and hand appearing too far from the actual or hands might interpenetrate the objects. To avoid these problems, we follow [17] to jointly optimize the physical hand and object poses by minimizing the ***Interaction Loss*** and the ***Collision Loss***.

Interaction Loss Due to estimation errors, hand poses and object poses can be distant from each other in the 3D space even though contact happens in reality. To diminish the distance, we minimize the interaction loss based on the Chamfer distance when contact happens. For every vertex within the hand mesh, the Chamfer distance function calculates the distance to the nearest point in the object mesh and subsequently aggregates the distances, as shown in Equation 1.

$$\begin{aligned} L_{Interaction} = & \frac{1}{|\mathbf{V}_{object}|} \sum_{x \in \mathbf{V}_{object}} \min_{y \in \mathbf{V}_{hand}} ||x - y||_2 \\ & + \frac{1}{|\mathbf{V}_{hand}|} \sum_{x \in \mathbf{V}_{hand}} \min_{y \in \mathbf{V}_{object}} ||x - y||_2 \end{aligned} \quad (1)$$

Collision Loss Object poses can interpenetrate hand poses, i.e. colliding and penetrating the hands. To resolve this collision issue, we penalize object vertices that are inside the hand using the collision loss function. A Signed Distance Field function (SDF) (Equation 2) is used to check if the object vertices are inside the hand.

$$\phi(\mathbf{v}) = -\min(SDF(v_x, v_y, v_z), 0) \quad (2)$$

If the cell is inside the hand mesh, ϕ takes positive values proproportional to the distance from the hand surface, and ϕ is 0 otherwise. The collision loss can be calculated as:

$$L_{collision} = \sum_{\mathbf{v} \in \mathbf{V}_{object}} \phi(\mathbf{v}) \quad (3)$$

The overall joint optimization function is then, where $\hat{\theta}$ is the optimized hand pose:

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^{45}}{\operatorname{argmin}}(L_{Interaction} + L_{collision}) \quad (4)$$

4.5.4 Contact Tracking. With hands and objects tracked, we calculate the hand-object contact using a similar approach described in [58]. By finding the nearest vertices on the object within a certain threshold for each vertex in the hand mesh, we identify the points of contact between the hand and the object. The histogram that is computed by counting the number of neighbors for each vertex of the MANO mesh can then be used to normalize and model the contact hotspots on the hand, which yields a more accurate representation of the contact points. The same process is repeated for the object mesh to generate a contact map on the surface of the object.

4.5.5 Mapping. The mapping process utilizes the information from the frame-by-frame tracking detailed in the previous subsection.

To map the object 6 DoF from the physical to the virtual, the user first aligns the objects' initial 6 DoF by moving and overlaying the virtual object to the physical one. We interpolate by [73] the shape from the physical object to the virtual object and store the interpolation information. Then, the translation and rotation of the physical object are measured frame-wisely and transform the virtual object in the 3D space. This allows the virtual object to match the position and orientation of the physical object to accurately transfer the contact points.

Once we transform the 6 DoF, we leverage the interpolation information to transfer the calculated contact points from the physical object to the virtual one every frame, following the methodology in [96].

After transferring the contact points, we eventually optimize the interaction between the virtual hand and the virtual object. In addition to the loss function in Equation 4, we also penalize the

Contact Loss by computing the Chamfer distance between the virtual hand and the set of mapped contact points C on the virtual objects as formulated in Equation 5.

$$L_{contact} = \frac{1}{|\mathbf{C}|} \sum_{x \in \mathbf{C}} \min_{y \in \mathbf{V}_{hand}} ||x - y||_2 + \frac{1}{|\mathbf{V}_{hand}|} \sum_{x \in \mathbf{V}_{hand}} \min_{y \in \mathbf{C}} ||x - y||_2 \tag{5}$$

The overall joint optimization function is the following, where $\hat{\theta}$ is the optimized hand pose interacting with the virtual object:

$$\hat{\theta} = \underset{\theta \in \mathbb{R}^{45}}{\operatorname{argmin}} (L_{Interaction} + L_{collision} + L_{contact}) \tag{6}$$

4.6 AR Interface

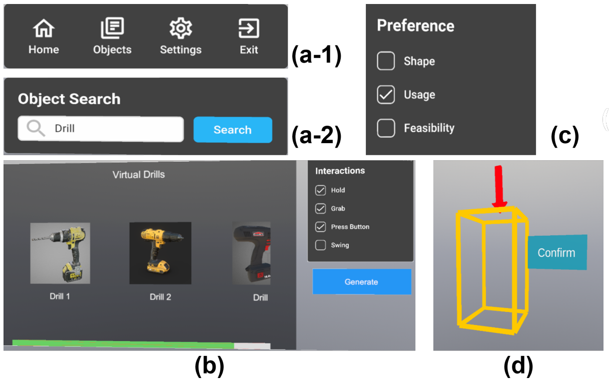


Figure 10: The AR interface of Ubi-TOUCH. (a-1) The main menu. The user can select to search for virtual objects, adjust some settings, return from mapping, or exit AR from mapping. (a-2) The search bar, where users just type the name of the virtual objects they would like to search. (b) The virtual object library. Where the user can select virtual objects by clicking on the intended snapshot on the left, select the interactions they would like to perform with this object, and eventually confirm by clicking the generate button. (c) The user can select the preference of the recommendation, by shape, usage, or feasibility of the objects. (d) All possible proxies will be marked with a 3D bounding box. The system will recommend the objects with the highest score per user preference, by indicating a red arrow above the bounding box.

We created an AR interface to incorporate the functionalities described in the previous sections. The AR interface has three modes. (1) **Search mode**, which allows the user to search the virtual object with keyboard inputs; (2) **Browse mode**, for visualizing the virtual objects and configuring the interaction to be applied to the object; and (3) **Scene mode**, for suggesting the potential tangible proxy to users and permitting them to choose the best fit based on their preference. The AR main menu is linked to the upper left of the user's field of view to facilitate mode switching. When the user activates the scan mode, a choice panel will appear on the top right of their sight, allowing them to select their preferred recommendation.

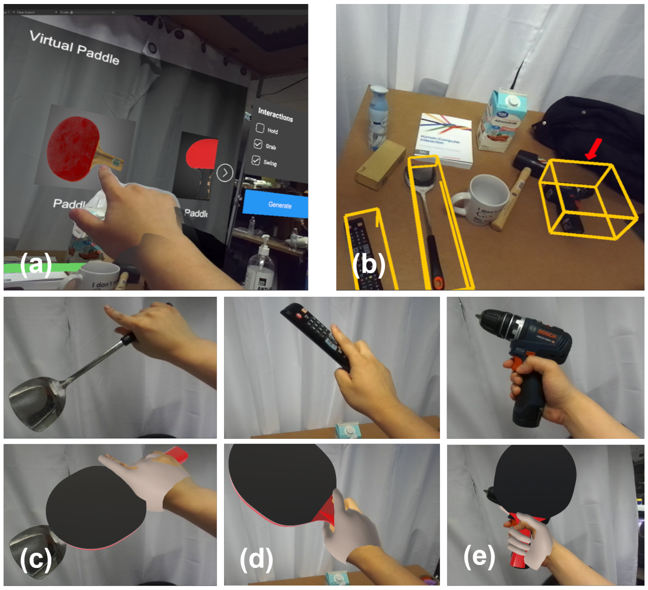


Figure 11: A table tennis practicing scenario. (a) The user selects a table tennis paddle from the library. (b) Ubi-TOUCH scans the vicinity of the user and marks the available proxies. The most preferred proxy is marked with a red arrow, in this case, a hand drill. (c,d,e) The user can grab the proxies while the as if grabbing the virtual paddle. The user will be shown the AR blending of the virtual hand and object. Then they can start playing with the virtual paddle either to practice alone or practice together in AR applications with the proxies.

As shown in Figure 10 a-2, users initiate the creation of the tangible proxy by entering the ***Search mode*** to search for the virtual object they intend to interact with. After entering the object's name into the system, the user can switch to ***Browse mode*** to view the available virtual objects to be used in the AR application. Then, the user needs to decide the target interaction using the panel on the left side of the library's interface.

After determining the target interactions and the virtual object, the user can now scan the environment to register the surrounding objects with Ubi-TOUCH. During this step, a preference menu will appear on the top left of their sight, allowing them to personalize the recommendations based on their object shape, usage, and feasibility choices. Then, Ubi-TOUCH computes the similarity between the targeted interactions and the available interactions for registered objects. The object with the highest similarity score is indicated with a bounding box and an arrow in red above it. If the users are satisfied with the recommendation, they can confirm the tangible proxy with the button close to the bounding box.

5 USE CASES

Given a target interaction with a virtual object, Ubi-TOUCH assists the users in locating the best object in their vicinity to interact with, maps the real-world hand-object interaction to the virtual interaction, and enables control over the virtual contents and IoTs within AR applications. We demonstrate four different use cases of Ubi-TOUCH in the following.

5.1 Ad Hoc Objects as Interaction Proxies

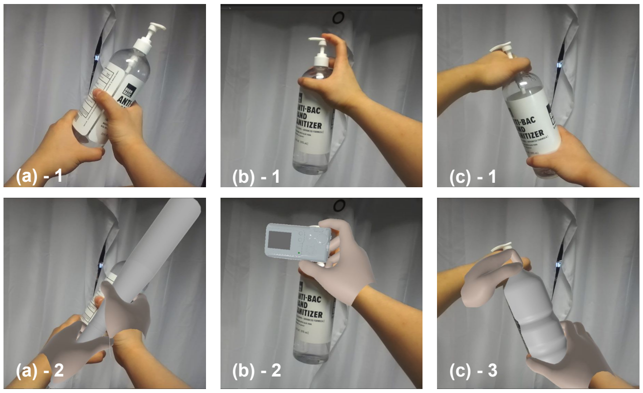


Figure 12: A demonstration of how interaction mapping is increasing the possibility of tangible proxies. (a)-1 A sanitizer dispenser can be grabbed, (b)-1 and pressed at the pump, (c)-1 and its cap can be screwed. The same interaction can be mapped into the virtual world by Ubi-TOUCH.(a)-2 A virtual baseball bat can be grabbed, (b)-2 a camera can be pressed at the shutter, and (c)-2 a bottle cap can be screwed in similar interactions with the dispenser.

Many objects can be interacted with in similar ways. Ubi-TOUCH takes advantage of the fact that the similarity between virtual interaction and physical interaction yields mental and physical consistency and immersiveness in user experience in AR applications [15, 61].

By utilizing hand-object interaction as the criterion, Ubi-TOUCH expands the range of possible proxies for tangible AR by diminishing the constraint of the physical object geometry without sacrificing consistency in user experience. As shown in Figure 11 (a), a user would like to practice table tennis with a virtual paddle by *grabbing* the handle and *swinging* the paddle. Given the target interactions *grabbing* and *swinging*, Ubi-TOUCH locates a remote, a screwdriver, and a spatula after scanning the users' vicinity (Figure 11 (b)). They can be grabbed and swung similarly to a paddle, despite the different geometry. Ubi-TOUCH recommends the interaction with those objects based on the similarity scores as shown in Figure 11 (b). Upon user selection, Ubi-TOUCH tracks the hand pose and the object 6 DoF in the physical world (Figure 11 (c,d,e)) and maps the hand-object interaction to the virtual world.

The possibility of tangible AR is enlarged by Ubi-TOUCH not only in the physical world but also in the virtual world. As shown in Figure 12, virtual interactions with diverse objects (*swinging* a baseball bat, *pressing* the shutter of a camera, and *screwing* the cap on a bottle) can all be respectively mapped into similar interactions with one object (*swinging* a dispenser, *pressing* the pump head of a dispenser, and *screwing* the cap on a dispenser).

5.2 Co-presenting Remote Hands-on Tutorial

Recent research has shown that AR can provide a more immersive and effective approach to hands-on training and education when combined with a sense of co-presence [6, 16, 67]. Ubi-TOUCH empowers such AR applications with more realistic hand-object interactions with haptic feedback.

In Figure 13, we showcase a one-on-one remote tutoring scenario. A *teacher* in the office tutoring a *learner* in the factory with the use of a hand drill. However, the *teacher* possesses a different hand drill (A) from that (B) of the *learner* (Figure 13 (b)-1). Ubi-TOUCH scans the vicinity in the office and suggests hand drill A to the *teacher* as the best-available proxy to interact with. To teach the *learner* how to *grab* and *hold* the hand drill as well as *press* the power button, the *teacher* demonstrates the interactions. Ubi-TOUCH captures the *teacher*'s interaction with hand drill A (Figure 13 (c)), creates the virtual counterpart of this interaction with hand drill B, and then displays in real-time the rendered virtual interaction to the *learner* (Figure 13 (d)). Despite the differences in object geometry and hand gesture, Ubi-TOUCH is able to map the *grabbing*, *holding*, and *pressing* interactions with hand drill A to corresponding interactions with hand drill B and provides accurate instruction to the *learner* as well as realistic haptic feedback to the *teacher*.

Considering a more challenging case, where the *teacher* does not possess any hand drill in the office, Ubi-TOUCH scans the vicinity and looks for the best-available object to *grab*, *hold*, and *press* like a hand drill. It eventually suggests the sprinkler (Figure 13 (b)-2). The *teacher* interacts with the sprinkler as a proxy. The target interactions are mapped to those with hand drill B and rendered in the *learner*'s display as instruction(Figure 13 (c)-2).

5.3 Tangible User Interface for Smart Homes

Recent development in the Internet of Things (IoT) has enabled the deployment of Smart Home devices and appliances that are interconnected through the IoT technology, enabling automation, remote control, and monitoring of household tasks and systems. Ubi-TOUCH can also be applied to prototype Tangible User Interface (TUI) in AR to control Smart Homes.

We demonstrate an example in Figure 14. Given any predefined interaction with a Smart Homes controller as the template, Ubi-TOUCH suggests all possible nearby objects that can be assigned the same functionality (buttons, sliders) as the virtual controller and can be interacted with similarly (Figure 14 (a,b)). Upon user selection, Ubi-TOUCH tracks the interactions with the selected object and overlays the virtual functionality onto the object (Figure 14 (a)-2,(b)-2). The user can *hold* the controller towards a Smart Home device *press* the virtual button by *pressing* on the designated part of the object to switch on and off the device in the room Figure 14 (b)-3. Meanwhile, the user can adjust the brightness of the light by *sliding* their fingers on the virtual slider while *holding* the controller towards the device. The same interactions can be mapped to different possible objects by Ubi-TOUCH as shown in Figure 14 (a)-3.

5.4 Interactive Tangible AR Game

Ubi-TOUCH can also benefit users with tangible controllers in various AR gaming scenarios.

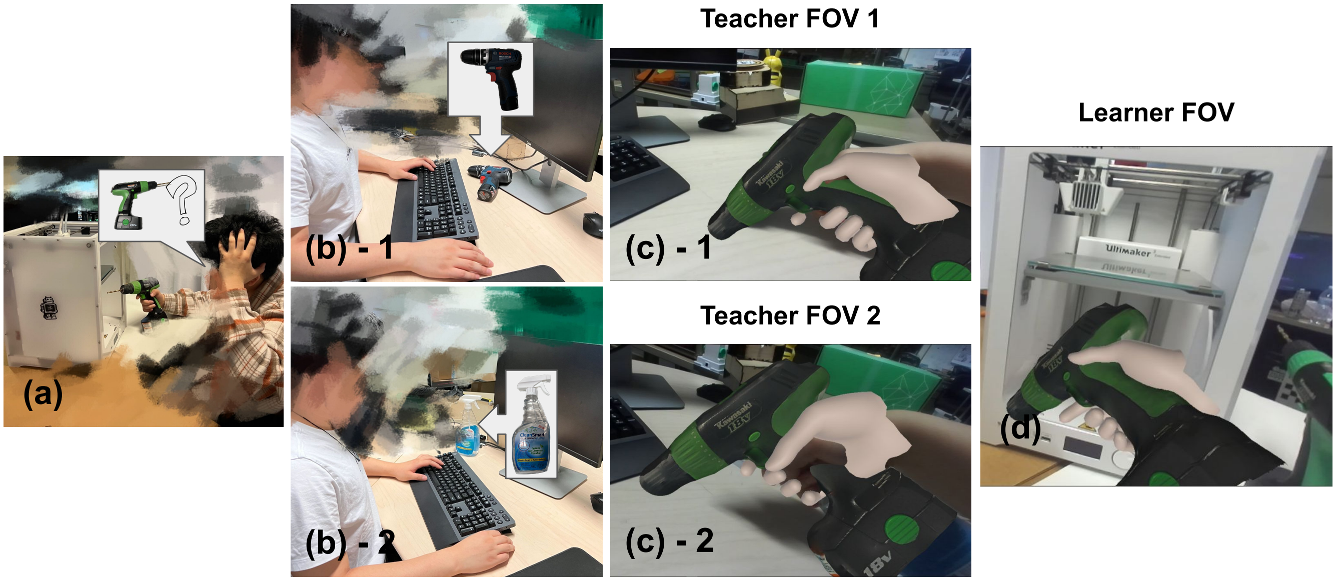


Figure 13: A Co-presenting Remote Hands-on Tutorial scenario. (a) A *learner* is having difficulties using a drill on the workbench. He called his *teacher* who is at the office with another drill (b)-1 or with only a sprayer (b)-2 nearby. Then the *teacher* utilizes Ubi-TOUCH to locate the proxy for the drill of the *learner* and map his interaction with this proxy ((c)-1,2). The same HOI is rendered to the *learner*'s HMD instructing him how to use this drill (d).

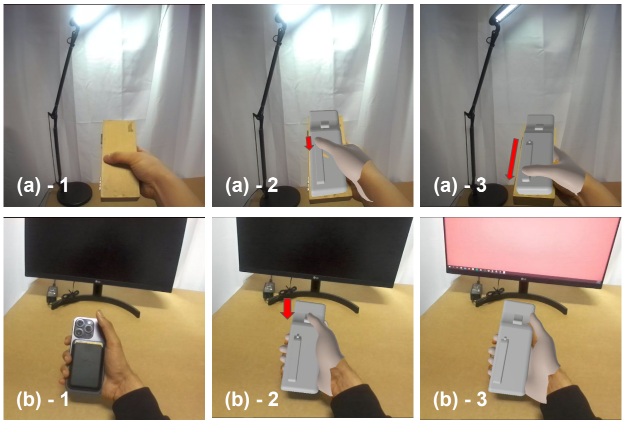


Figure 14: A demonstration of Ubi-TOUCH as a tangible user interface. (a)-1 and (a)-2 the user uses the glass box as the virtual remote. The user slide down the virtual slider on the remote, and (a)-3 the brightness of the light goes down. Similarly, in (b)-1, the user uses the back of a cell phone as the remote. The user presses the virtual button (b)-2 and switches on the monitor (b)-3.

As shown in Figure 15, the user wants to play a balloon-shooting AR game and seeks a proxy for a Nerf Gun (Figure 15 (a)). Ubi-TOUCH scans for objects that can be *grabbed* and *pressed* (pulling the trigger) and suggests a sprayer and a drill (Figure 15 (a)). Both of them can be interacted with as a proxy for the Nerf Gun. By blending the consistent HOI into the physical world, Ubi-TOUCH enables the users to immersively interact with the virtual objects with proximate haptic feedback. When the trigger on the drill is pulled by the user, the same virtual interaction will be mapped to, rendered, and blended into the display of the user (Figure 15 (c)). The user can aim with the virtual Nerf Gun by moving the drill, and shoot the virtual balloons as shown in Figure 15 (c) by physically pulling the trigger.

6 SOFTWARE AND HARDWARE IMPLEMENTATION

We implemented Ubi-TOUCH using a customized Oculus Quest 2 [70] HMD with a ZED [86] camera with additional depth data for AR pass-through experiences. A Type- C cable connected Oculus with ZED camera to a local PC (Intel core i7-9700K CPU, 26 GHz, 64 GB RAM). The Ubi-TOUCH interface is developed on Unity 3D (2021.3.8f1). The AR view rendering is implemented using the ZED-unity plugin. To allow interaction of the physical hand with the virtual object mentioned in section 4.1, an inbuilt hand interaction plugin from Oculus was used. The ZED mini captures the RGB frames and depth images and displays them using the HMD device. During the object and hand pose tracking pipeline we adopt a resolution of 1280 x 720 for both RGB color image and a depth image. The algorithms were implemented on 2 Nvidia GeForce GTX 2080 Ti GPUs in the PC. The frames were captured at 15 FPS and were rendered and displayed after processing through the HMD device at 5 fps.

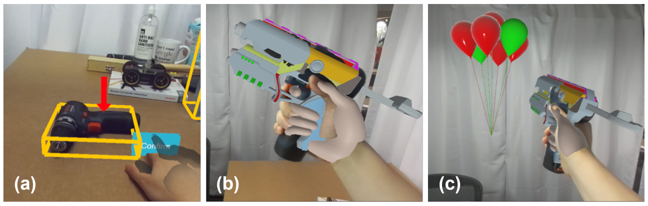


Figure 15: A demonstration of an AR shooting game with Ubi-TOUCH. The user wants a proxy for a virtual Nerf Gun and is suggested two proxies, a sprayer and a drill in the environment (a). The user's interaction with the drill is then mapped to the interaction with the Nerf Gun (b). The user then uses the drill to aim at some virtual balloons and shoot (c).

7 PRELIMNIARY SYSTEM FOR EVALUATING HAND-OBJECT INTERACTION

In real-world scenarios, tracking both hands and objects is difficult with RGB images. Although there are many solutions to track both objects and hands using additional hardware capabilities hardware such as mocap and antilatency, their setup requires extra cost and limits the use case of the tracking. Due to advances in the current computer vision area, algorithms can track both hand and object from single RGB images but the accuracy is not at par for novel objects not seen during the training. Thus in this section, we evaluate the effectiveness of our hand and object pose tracking algorithm, which is important for the accurate mapping of hand-object interactions.

7.1 Collection

To evaluate our single view algorithm as mentioned in 4.1, similar to [58], we follow auto annotated data collection strategy using a 5-camera hardware system as shown in Figure 6 to generate the auto-annotated training data for the hand object pose tracking and interactions. We collected our own training data instead of evaluating other computer benchmark datasets to get a better interpretation of the performances of algorithms in practice (objects in use). 10 participants were recruited to collect the entire dataset. We follow the standard way to make our dataset with varying factors such as occlusion and lighting conditions. The whole dataset contains interaction with the object with only one hand (Right). We collected 1000 videos for 10 classes of interactions.

7.2 Verification

We verify the accuracy of the hand-object pose annotations from our hardware setup on 500 images on 5 different camera views from our data set similar to H20 [58]. Three experts with prior data annotation experience were recruited to manually annotate 500 images for both hand and object. For hands, experts annotate 2D key points (joints) on the hands which we triangulate to get 3D annotations. For objects, experts use the 6 Dof annotation tool [1]. We compute the mean per joint point error (in cm) over 21 joints(MPJPE) following [100] to check the annotation quality for hands. For the object, The following metrics were used: R_{err}: mean orientation error in degrees. T_{err}: mean translation error in centimeters. The lower is better for orientation errors and translation errors.

For Hand, the average MPJPE is 1.1 cm with a standard deviation of 0.4, and for objects, T_{err} and R_{err} were 1.5 cm and 1.7°. For both hand and object the error range lies within a range of 1 cm and 1° which shows high-ground truth annotations for our dataset [58]. The whole dataset consists of 10 interaction verb classes with 10 objects.

7.3 Procedure

To evaluate the algorithm performance, we selected 5 hand-object interactions with 5 objects, i.e. a sprayer, a drill gun, a cup, a bottle, and a hammer from our dataset. To test our hand-object pose tracking model, we compare the accuracy with ground truth from our dataset for both hand and object. We further compare the accuracy of our model with BundleTrack [93], one of the state-of-the-art algorithms for object tracking in the wild(objects). For the hands, we use HandOccNet [72] for comparison.

7.4 Results and Discussion

The results are summarised in the table Table 1 and 2. The algorithm is able to track the object pose of a drill and book with the highest accuracy among all 5 and the lowest is the hammer. Objects with rich textures, such as those with clear edges, corners, or distinctive patterns, tend to have more features that can be tracked, which makes it easier to track. Bottle performance was bad because of the textureless surface. On the hand, we achieve the best performance with a hammer. The screwdriver is the least occlusion of the hand among other objects. We discuss other failure cases of the algorithm in section 9. Overall the results show comparable performance to the state of the art.

8 USER STUDY

We conducted a user study to evaluate the accuracy of the recommendation module, the mapping of hand-object interactions, and the overall system usability. We invited twelve participants (three female; nine male) from a technical university's graduate and undergraduate program. 10 of the users had experience with AR/VR applications using tablets, smartphones, and head-mounted devices. 2 out of 12 had a basic understanding of AR/VR concepts. We did not invite any users with prior AR/VR application programming experience since Ubi-TOUCH is designed to provide a tangible virtual hand-object interactions experience to non-expert AR consumers. None of the users have prior user study experience specifically for AR/VR applications. The user's age ranges from 18 to 29 with a mean of 24.5 years. The entire study took 1.5 hours and each user was paid a 15 USD e-gift card. The study was conducted in a 5m x 5m indoor environment and screen recorded for post-analysis. Upon users' arrival, an explanation of the study was provided, followed by a signature on the consent form. After that, we explained our system and let users understand the entire system workflow and system UI. Before the user's study officially started, we provided some time for first-time users experiencing Oculus Quest to get comfortable. Considering counter balancing we divided 12 participants into a group of 2 with 6 users each. After completing the session, a System Usability Test (SUS) [14] and a 5-scale Likert-type

Table 1: The object pose testing results on the collected dataset and comparison with the benchmark with 5 objects.

Method	bottle	drill	cup	book	hammer
Rotation Error R_{err} (degree) (lower the better)					
BundleTrack [93]	13.16	5.23	7.41	4.2	15.6
Ours	11.5	5.1	9.5	3.9	14.1
Translation Error T_{err} (cm) (lower the better)					
BundleTrack [93]	4.16	2.58	3.9	2.13	5.4
Ours	4.02	2.61	3.87	1.95	3.9

Table 2: The hand testing results on the collected dataset and comparison with the benchmark.

Model	MPJPE (cm) (lower the better)
HandOccNet [72]	2.53
Ours	2.39

questionnaire were administered to the users for the usability of Ubi-TOUCH. We also conducted post-session conversation-type interviews to get subjective feedback.

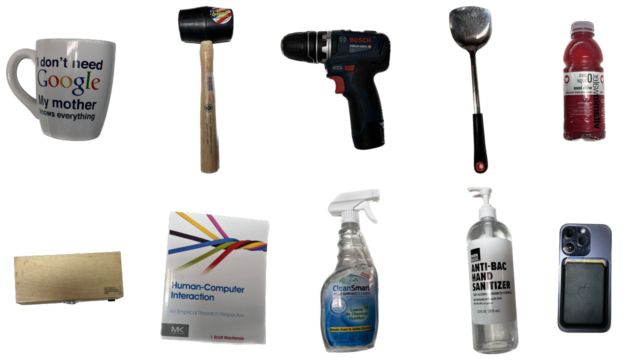
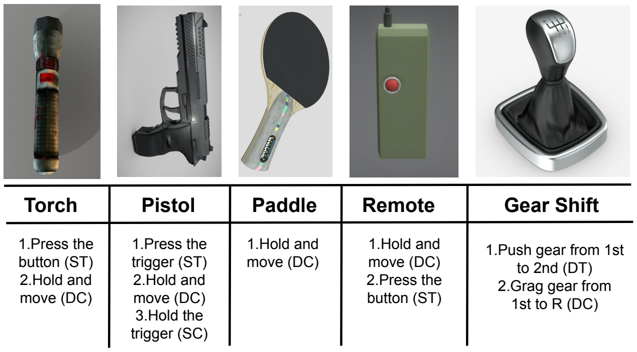


Figure 16: Ten objects used as a proxy object for the user study. Top Left to Right: Mug, Hammer, Drill gun, Spatula, and Bottle. Bottom Left to Right: Box, Book, Sprinkler, handwash, and Mobile

8.1 Procedure

We evaluated the performance of our interface and let the users experience virtual object interactions with 5 different virtual objects as shown in Figure 17. In the user testing session, participants were tasked with completing ten interactions with 5 different virtual objects using physical proxy objects. To accomplish this, ten physical objects (Figure 16) were provided as proxy options for interacting with the virtual objects. 1) Moving the virtual remote (Static Transient) and pressing a button(Dynamic Continuous), 2) Grabbing and moving the torch (Dynamic Continuous) and pressing a button (Static Transient), 3) playing with a ping pong paddle (Dynamic continuous), 4) playing with a pistol specifically grabbing and moving(Dynamic Continuous) pressing the trigger(Static Transient), holding the trigger (Static Continuous), and 5) pushing a gear shift from first gear to second gear (Dynamic Transient) and moving from first to Reverse (Dynamic Continuous).



Torch	Pistol	Paddle	Remote	Gear Shift
1.Press the button (ST) 2.Hold and move (DC)	1.Press the trigger (ST) 2.Hold and move (DC) 3.Hold the trigger (SC)	1.Hold and move (DC)	1.Hold and move (DC) 2.Press the button (ST)	1.Push gear from 1st to 2nd (DT) 2.Grag gear from 1st to R (DC)

Figure 17: User Study: 5 virtual objects covering 4 Interactions which Static Transient(ST), Static Continuous(SC), Dynamic Transient(DT), and Dynamic Continuous(DC)

Users were shown the virtual model and were asked to select the type(s) of interactions they want to perform with the virtual object. For the first 5 interactions, the first group was asked to imagine physical objects among 10 objects present in the vicinity that can match the virtual model to perform interactions. Next, users were asked to point out the objects they imagined which can be used as a real proxy. We recorded the possible selections users made. For the next 5 interactions system recommends (more than one) proxy object that can be used by the user. For the second group, the system recommended the first five interactions, and then for the later 5 they imagined the interactions, and then the system recommended. Then, the user aligns the virtual object with the real proxy and then performs the interaction(s) with the real object.

8.1.1 Recommendation. We qualitatively evaluate our recommendation module using Likert Scale Questionnaire. The results are shown in Figure 18. Many users acknowledge the comprehensive proxy object recommendation and the accuracy of the recommendations ((Q12: AVG=3.8, SD=1.3) and (Q13: AVG=4, SD=0.9)). *"When I got the recommendation from the system, I was amazed to get a physical object that I imagined (P11)."* . Most of the users agreed that our recommendation provides more reasonable proxies and acknowledges wider options for proxy recommendations(Q14: AVG=4.1, SD=1.1).*"It is awesome that I can get more options for interacting with virtual objects (P7)."*.

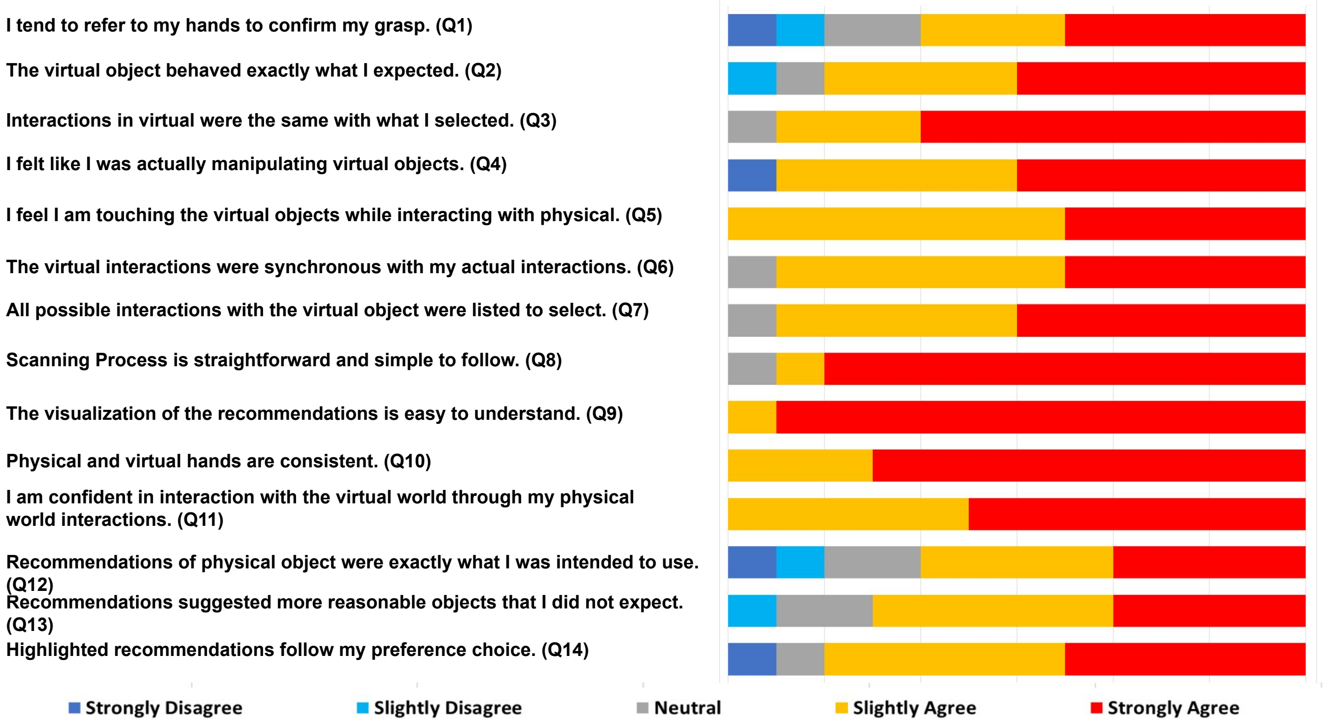


Figure 18: Likert-type questionnaire results

8.1.2 Mapping of Hand object interaction. We qualitatively evaluate hand-object interaction mapping in interactions performed by the user during the study. We evaluated using Likert scale results shown in Figure 18. Users preferred the visualization of virtual hand-object together(Q1: AVG=3.8, SD=1.3). *"I think the virtual hand guided me to play with the virtual object (P5).*. Most of the users found good interaction transferring from physical to virtual(Q2: AVG=4.3, SD=0.9). *"Oh this is so cool, the virtual object is exactly moving the way I want." (P8).* Many users acknowledged that the virtual interactions were exactly the same as what they intended to perform (Q3: AVG=4.6, SD=0.7). *"I must say the interactions look so real and are exactly what I wanted to do (P1).".* Most users agree that object tracking is satisfactory (Q4: AVG=4.3, SD=1.1). *"The virtual object is rotating in the way of the object present in my hand (P5)".* Users were able to get haptic for the virtual interactions (Q5: AVG=4.4, SD=0.5). *"I can get the same feeling as if I am touching and grabbing the physical objects(P9)".* Finally, many users welcome the synchronous interactions between both real and virtual (Q6: AVG=4.3, SD=0.6). *"It is good that your interaction transferring is synchronous and I can follow real virtual interactions both (P11)".*

8.1.3 Overall System Usability. The overall system Likert results collected are shown in Figure 18. In general, users prefer using everyday objects around them as proxies for interacting with virtual objects (Q11: AVG=4.6, SD=0.5). *"In my opinion, I can actually feel the virtual object and the interactions were similar to the virtual object using real (P8)".* Though we did not explicitly ask users, still some of them found the system fun to use. *"The system was easy to follow and fun to use (P7)".* We asked users about each module in our interface. Users acknowledge the comprehensiveness of possible interactions that can be performed with the virtual object(Q7: AVG=4.4, SD=0.6). *"All the interactions were present that I can think of, especially with the gearbox object. (P2)".* Regarding the environment scanning process, many users provided us with satisfactory remarks (Q8: AVG=4.8, SD=0.6). *"The procedure of scanning was straightforward, and amazed that we can scan the environment easily (P12)".* Most of the users favored the way recommendations were presented to them (Q9: AVG=4.9, SD=0.3). *"I was able to understand the objects recommended by the system and scores were helpful(P6)".*

Many users found a clear separation between the physical hand-object interaction and virtual hand-object interaction(Q10: AVG=4.8, SD=0.5). *"Virtual hand and object were clearly visible and the hand shape makes sense (P4)".* For the system usability, the users reported an M = 86.45 out of 100 and SD = 13.03 SUS. This score is promising and indicates the high usability of the system.

9 LIMITATIONS AND FUTURE WORK

9.1 Visual Distraction from Physical Hand-object Interaction

The core motivation of our work is to provide more options for tangible proxies for AR while maintaining consistency in users' HOI experience. To address this point, we propose an algorithm to render a consistent Hand-object interaction and overlay it in AR. However, users report that they can be distracted by both Physical

HOI and virtual HOI appearing. *"I saw my own hand and the virtual hand. I did not know which one I should see(P5).".* A similar problem is that the physical object still appears in the environment and there may be chances that the virtual hands will penetrate the physical objects. *"It was confusing cause my virtual hand was in the bottle and it was blocked by it. I could not see my fingers"(P8).*

This problem of occlusion has been addressed in several prior works [48, 49, 56, 63]. In our scope, we seek to diminish the visual interference of the physical HOI to the virtual HOI. We envision in the future iteration of our system, we can blur the physical HOI visuals on the AR device, or remove the physical HOI, interpolate, and render the background. This can be accomplished by utilizing multiple cameras and reconstructing the point cloud. Incorporating known physical HOI information detected as described in our methodology, the corresponding point cloud can be preprocessed in order to eliminate the visual distraction from the physical HOI. An easy implementation of this method can be overlaying the virtual components onto the projection of the physical hands and objects to hide them.

9.2 Extreme Mapping Cases

Ubi-TOUCH aims to provide a wider range of choices for tangible proxies. However, there are cases where the objects are not enough in the vicinity of the users and where the users' selections override the system's criteria. *"I just want to see what happens if I use a bottle as a remote(P3).".* When an object is selected as the proxy of a distinctive target, which does not resemble in shape and cannot be interacted with in the same way, contact point mapping in our algorithm will take priority, resulting in implausible interaction mappings. *"Then, it put the remote button onto the tip of the bottle, which I cannot reach with any of my fingers when grabbing the bottle(P12)"*

When the degree of freedom in the physical world is limited as stated above, we can approach the mapping problem from the virtual world. Given a few physical objects which cannot be mapped to virtual, a possible solution is to fit the virtual object into the physical objects, to which the users have access. For example, a physical bottle is hard to be mapped into a virtual remote, and the bottle is the only object that the user has in his vicinity. Under such circumstances, we can reverse our contact mapping process, using the physical object shape and the physical hand gesture as the constraints to fit the virtual object into the shape of the physical object. In our example, this solution will yield a bottle-shaped virtual remote, where the buttons can be reached and pressed easily with the hand holding it exactly like holding a bottle.

To this end, we envision future work to consider the high degree of freedom in the virtual world. A reversed methodology (in terms of ours) can be applied to transform virtual objects into any physical proxy the user has in hand. Reasonable mapping is expected to satisfy requirements such as ergonomics, consistency in the functionality of the virtual object, and user experience.

9.3 Physical Properties of the Objects

Being a tangible proxy does not mean that the objects share identical haptic feedback. E.g., Touching an iron ball is different from touching a basketball. Users report that the material of the proxy mismatches with the target object, creating a gap between the physical world and the virtual world. *"The material of proxy cannot perfectly simulate the target. They just feel different. (P3)"* The material of an object is part of its physical Properties. Other physical Properties also play significant roles in matching the real-world HOI with the virtual ones, such as the mass, density, temperature, surface, and shape of the objects. *"It tells me to swing a hammer like swinging a ping-pong paddle, which is not realistic. The head of the hammer is heavy, but the paddle is not like this. (P5)"* To decide whether an object is a good fit, more knowledge is required regarding the physical attributes of this object. The mapping between interactions can be further optimized given the material knowledge of the objects and the physics of the interactions themselves [3, 8, 97]. We hope future research on tangible proxies for immersive technology can build convincing taxonomy of the physical attributes to provide a more consistent, realistic, and safer immersive experience.

9.4 Interaction Knowledge-base in the Wild

Some interactions are not covered in the database. *"I think I can use the handle of the cup as a pistol. But this was not suggested by the system. (P9)"* We collect the interaction knowledge from existing datasets, where the hand-object interactions are normal. Undeniably, there are more interactions that we do not normally perform, such as grabbing a cup by pinching the handle. However, we envision that such knowledge can unexpectedly help to map in extreme cases. Creative knowledge about how an object can be interacted with differently from its designated purposes can easily come from users. Utilizing such knowledge can further diminish the constraints on the choice of proxy objects by enabling novel affordances of the objects and distinctive gestures to interact with the objects. With this, we can discover a larger distribution of HOI. We envision follow-up work with a collective methodology to capture the users' novel interactions with objects in the wild.

9.5 Software and Hardware Constraints

We have demonstrated the ability of our system in the real-time mapping of hand-object interactions from physical to virtual. Although the system performs in real-time, some of the users mention more natural like rapid interactions than controlled movement. We used two GPUs with good computation capacity but still because computational processing time limitation for algorithms is way more than performing 30 fps or more in real-time. Further because of the static hardware, the system restricts users' ability to use it in other environments such that in the kitchen, factories, or even outside on the field while playing."The application would be more useful if I can freely move in the room and perform more interactions(P6)". The tracking and mapping component is the most computationally time-consuming because of object tracking (55 ms) and hand pose optimization (65 ms). The computational problem can be solved in the future by utilizing cloud services for data transferring and computation. Further, parallel programming for object 6 DoF tracking and usage of better GPUs (high computational) can contribute to better time performance of the system. Also, the reusing of object geometric features can reduce the time cost in the registration and recommendation modules. With the limited computation bandwidth, we still were able to reach satisfactory performance for hand and object pose tracking without additional sensors. However, there are still some inaccurate predictions with the system in cases of high occlusions and complex backgrounds. One possible solution could be to include additional sensors for tracking hand and object pose. Nevertheless, additional hardware will restrict mobility and constraints to a wider range of tasks and the environment's usability. We envision in the future that lighter hardware sensors and improved algorithms can solve both hand and object pose tracking in the wild.

9.6 Safety

What may be of limited exposure in our use case and user study is the safety concerns of the methodology of our system. In any case, we do not anticipate the users to use any physical objects that pose a threat or danger in any form to any personnel. To address this safety concern, we envision two major aspects to avoid the use of harm. First of all, any virtual objects that poses danger in the virtual world will not be mapped to its physical counterpart but to more constrained proxies. For instance, if a user is seeking proxies of a virtual knife, any physical knife or edgy object will not be considered. Secondly, when collecting the database for physical objects as well as when registering objects in the vicinity, we should exclude the objects with potentially harmful geometric attributes such as sharp edges and tips in any use case.

10 CONCLUSION

In this work, we present UBI-TOUCH, an AR system that enables users to use tangible physical objects for interacting with virtual objects. UBI- TOUCH provides haptic feedback to the user by using their own personalized objects while interacting with the virtual object. By blending the virtual HOI into the physical world, Ubi-TOUCH makes visible the consistent rendering of the virtual interaction to provide the user with an immersive experience. We first discuss a taxonomy for broadly classifying hand-object interactions based on hand-object touch time and object position. Following this taxonomy, we describe components of Hand object interactions that can be used to map real and virtual interactions. Then we present an overall workflow of the system and discuss four major modules: virtual object interaction type selection, scanning the environment, registering the objects to the system, proxy object recommendations, and then finally mapping of real hand object interaction to the virtual hand object interactions. To explore the scalability of Ubi-TOUCH, we demonstrate four different application scenarios which are: Ad hoc objects as interaction proxies, smart homes, Remote tutoring, and AR game. Through the user study, we first evaluated our hand-object pose tracking performance which demonstrated that UBI-TOUCH can accurately track the pose of the hand and object. Then we proved the usability of our system and its utility. Ubi-TOUCH was tested with 12 users and received positive feedback. Thus, we believe Ubi-TOUCH provides tangible feedback for the application that involves virtual hand-object interactions. We anticipate that this research paper will open up novel opportunities for the development of efficient remote tutoring systems and AR games with haptic feedback, thereby reaching a wider audience and making these technologies more accessible and engaging for diverse users.

ACKNOWLEDGMENTS

We wish to thank all the reviewers for their invaluable feedback. We thank Hyungjun Doh for his help in building the database. This work is partially supported by the NSF under the Future of Work at the Human-Technology Frontier (FW-HTF) 1839971. We also acknowledge the Feddersen Distinguished Professorship Funds and a gift from Thomas J. Malott. Any opinions, findings, and conclusions expressed in this material are those of the authors and do not necessarily reflect the views of the funding agency.

REFERENCES

[1] 6D-PAT. 2022. 6D. Retrieved April, 2020 from https://github.com/florianblume/6d-pat

[2] Parastoo Abtahi, Benoit Landry, Jackie Yang, Marco Pavone, Sean Follmer, and James A Landay. 2019. Beyond the force: Using quadcopters to appropriate objects and the environment for haptics in virtual reality. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[3] Jacopo Aleotti, Francesco Denaro, and Stefano Caselli. 2010. Object manipulation in visuo-haptic augmented reality with physics-based animation. In *19th International Symposium in Robot and Human Interactive Communication*. IEEE, 38–43.

[4] Jatin Arora, Aryan Saini, Nirmita Mehra, Varnit Jain, Shwetank Shrey, and Aman Parnami. 2019. Virtualbricks: Exploring a scalable, modular toolkit for enabling physical manipulation in vr. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[5] K. S. Arun, T. S. Huang, and S. D. Blostein. 1987. Least-Squares Fitting of Two 3-D Point Sets. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-9, 5 (1987), 698–700. https://doi.org/10.1109/TPAMI.1987.4767965

[6] Huidong Bai, Prasanth Sasikumar, Jing Yang, and Mark Billinghurst. 2020. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376550

[7] Saskia Bakker, Alissa N Antle, and Elise Van Den Hoven. 2012. Embodied metaphors in tangible interaction design. *Personal and Ubiquitous Computing* 16 (2012), 433–449.

[8] David Beaney and Brian Mac Namee. 2009. Forked! A demonstration of physics realism in augmented reality. In *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. IEEE, 171–172.

[9] Oriel Bergig, Eyal Soreq, Nate Hagbi, Kirill Pevzner, Nati Levi, Shoham Blau, Yulia Smelansky, and Jihad El-Sana. 2011. Out of the cube: augmented Rubik's cube. *International Journal of Computer Games Technology* 2011 (2011), 2–2.

[10] Carlos Bermejo and Pan Hui. 2021. A survey on haptic technologies for mobile augmented reality. *ACM Computing Surveys (CSUR)* 54, 9 (2021), 1–35.

[11] Mark Billinghurst, Hirokazu Kato, Ivan Poupyrev, et al. 2008. Tangible augmented reality. *Acm siggraph asia* 7, 2 (2008), 1–10.

[12] 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)*. IEEE, 778–787.

[13] Samarth Brahmbhatt, Cusuh Ham, Charles C Kemp, and James Hays. 2019. Contactdb: Analyzing and predicting grasp contact via thermal imaging. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 8709–8719.

[14] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[15] Keith R Bujak, Iulian Radu, Richard Catrambone, Blair MacIntyre, Ruby Zheng, and Gary Golubski. 2013. A psychological perspective on augmented reality in the mathematics classroom. *Computers & Education* 68 (2013), 536–544.

[16] Yuanzhi Cao, Xun Qian, Tianyi Wang, Rachel Lee, Ke Huo, and Karthik Ramani. 2020. An Exploratory Study of Augmented Reality Presence for Tutoring Machine Tasks. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376688

[17] Zhe Cao, Ilija Radosavovic, Angjoo Kanazawa, and Jitendra Malik. 2021. Reconstructing hand-object interactions in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 12417–12426.

[18] R. Charles, H. Su, M. Kaichun, and L. J. Guibas. 2017. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 77–85. https://doi.org/10.1109/CVPR.2017.16

[19] Yang-Sheng Chen, Ping-Hsuan Han, Jui-Chun Hsiao, Kong-Chang Lee, Chiao-En Hsieh, Kuan-Yin Lu, Chien-Hsing Chou, and Yi-Ping Hung. 2016. Soes: Attachable augmented haptic on gaming controller for immersive interaction.

In *Adjunct Proceedings of the 29th Annual ACM Symposium on User Interface Software and Technology.* 71–72.

[20] Lung-Pan Cheng, Eyal Ofek, Christian Holz, Hrvoje Benko, and Andrew D Wilson. 2017. Sparse haptic proxy: Touch feedback in virtual environments using a general passive prop. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems.* 3718–3728.

[21] Jaeyong Chung, Hwan-Jin Yoon, and Henry J Gardner. 2010. Analysis of break in presence during game play using a linear mixed model. *ETRI journal* 32, 5 (2010), 687–694.

[22] Florian Daiber, Donald Degraen, André Zenner, Tanja Döring, Frank Steinicke, Oscar Javier Ariza Nunez, and Adalberto L Simeone. 2021. Everyday Proxy Objects for Virtual Reality. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–6.

[23] Dragoş Datcu, Stephan Lukosch, and Frances Brazier. 2015. On the usability and effectiveness of different interaction types in augmented reality. *International Journal of Human-Computer Interaction* 31, 3 (2015), 193–209.

[24] Xavier de Tinguy, Claudio Pacchierotti, Mathieu Emily, Mathilde Chevalier, Aurélie Guignardat, Morgan Guillaudeux, Chloé Six, Anatole Lécuyer, and Maud Marchal. 2019. How different tangible and virtual objects can be while still feeling the same?. In *2019 IEEE World Haptics Conference (WHC).* IEEE, 580–585.

[25] Xavier de Tinguy, Claudio Pacchierotti, Maud Marchal, and Anatole Lecuyer. 2019. Toward universal tangible objects: Optimizing haptic pinching sensations in 3d interaction. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR).* IEEE, 321–330.

[26] 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.* 1–4.

[27] 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.* 1–4.

[28] Tim Düwel, Martin Feick, and Antonio Krüger. 2021. Considering Interaction Types for Geometric Primitive Matching. (2021).

[29] Benjamin Eckstein, Eva Krapp, Anne Elsässer, and Birgit Lugrin. 2019. Smart substitutional reality: Integrating the smart home into virtual reality. *Entertainment Computing* 31 (2019), 100306.

[30] Cathy Mengying Fang and Chris Harrison. 2021. Retargeted self-haptics for increased immersion in vr without instrumentation. In *The 34th Annual ACM Symposium on User Interface Software and Technology.* 1109–1121.

[31] Cathy Mengying Fang, Ryo Suzuki, and Daniel Leithinger. 2023. VR Haptics at Home: Repurposing Everyday Objects and Environment for Casual and On-Demand VR Haptic Experiences. *arXiv preprint arXiv:2303.07948* (2023).

[32] Jacqui Fashimpaur, Kenrick Kin, and Matt Longest. 2020. Pinchtype: Text entry for virtual and augmented reality using comfortable thumb to fingertip pinches. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems.* 1–7.

[33] 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).* IEEE, 195–206.

[34] Jingyu Shi Xun Qian Tianyi Wang Karthik Ramani Fengming He, Xiyun Hu. 2023. Ubi Edge: Authoring Edge-Based Opportunistic Tangible User Interfaces in Augmented Reality. https://engineering.purdue.edu/cdesign/wp/ubi-edge-authoring-edge-based-opportunistic-tangible-user-interfaces-in-augmented-reality/. Online; accessed April 5 2023, to be published.

[35] Sean Follmer, Daniel Leithinger, Alex Olwal, Akimitsu Hogge, and Hiroshi Ishii. 2013. inFORM: dynamic physical affordances and constraints through shape and object actuation.. In *Uist*, Vol. 13. Citeseer, 2501–988.

[36] Claude Ghaoui. 2005. *Encyclopedia of human computer interaction.* IGI global.

[37] James J Gibson. 1977. The theory of affordances. *Hilldale, USA* 1, 2 (1977), 67–82.

[38] Raphael Grasset, Andreas Duenser, Hartmut Seichter, and Mark Billinghurst. 2007. The mixed reality book: a new multimedia reading experience. In *CHI'07 extended abstracts on Human factors in computing systems.* 1953–1958.

[39] 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).* IEEE, 842–843.

[40] Aakar Gupta, Bo Rui Lin, Siyi Ji, Arjav Patel, and Daniel Vogel. 2020. Replicate and reuse: Tangible interaction design for digitally-augmented physical media objects. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–12.

[41] 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.* 211–218.

[42] 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.* 1957–1967.

[43] Eva Hornecker and Jacob Buur. 2006. Getting a grip on tangible interaction: a framework on physical space and social interaction. In *Proceedings of the SIGCHI conference on Human Factors in computing systems.* 437–446.

[44] Wan-Ting Hsu and I-Chen Lin. 2021. Associating Real Objects with Virtual Models for VR Interaction. In *SIGGRAPH Asia 2021 Posters.* 1–2.

[45] Gaoping Huang, Xun Qian, Tianyi Wang, Fagun Patel, Maitreya Sreeram, Yuanzhi Cao, Karthik Ramani, and Alexander J Quinn. 2021. Adaptutar: An adaptive tutoring system for machine tasks in augmented reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–15.

[46] 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.* 234–241.

[47] Seokhee Jeon and Seungmoon Choi. 2009. Haptic augmented reality: Taxonomy and an example of stiffness modulation. *Presence* 18, 5 (2009), 387–408.

[48] Denis Kalkofen, Erick Mendez, and Dieter Schmalstieg. 2007. Interactive focus and context visualization for augmented reality. In *2007 6th IEEE and ACM International Symposium on Mixed and Augmented Reality.* IEEE, 191–201.

[49] Denis Kalkofen, Erick Mendez, and Dieter Schmalstieg. 2008. Comprehensible visualization for augmented reality. *IEEE transactions on visualization and computer graphics* 15, 2 (2008), 193–204.

[50] Byeongkeun Kang, Kar-Han Tan, Nan Jiang, Hung-Shuo Tai, Daniel Tretter, and Truong Nguyen. 2017. Hand segmentation for hand-object interaction from depth map. In *2017 IEEE Global Conference on Signal and Information Processing (GlobalSIP).* IEEE, 259–263.

[51] Hirokazu Kato, Mark Billinghurst, Ivan Poupyrev, Kenji Imamoto, and Keihachiro Tachibana. 2000. Virtual object manipulation on a table-top AR environment. In *Proceedings IEEE and ACM International Symposium on Augmented Reality (ISAR 2000).* Ieee, 111–119.

[52] Hanseob Kim, Myungho Lee, Gerard J Kim, and Jae-In Hwang. 2021. The impacts of visual effects on user perception with a virtual human in augmented reality conflict situations. *IEEE Access* 9 (2021), 35300–35312.

[53] Kangsoo Kim, Gerd Bruder, and Greg Welch. 2017. Exploring the effects of observed physicality conflicts on real-virtual human interaction in augmented reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology.* 1–7.

[54] Benjamin Knoerlein, Gábor Székely, and Matthias Harders. 2007. Visuo-haptic collaborative augmented reality ping-pong. In *Proceedings of the international conference on Advances in computer entertainment technology.* 91–94.

[55] Thomas Kosch and Albrecht Schmidt. 2020. Enabling Tangible Interaction through Detection and Augmentation of Everyday Objects. *arXiv preprint arXiv:2012.10904* (2020).

[56] Ernst Kruijff, J Edward Swan, and Steven Feiner. 2010. Perceptual issues in augmented reality revisited. In *2010 IEEE International Symposium on Mixed and Augmented Reality.* IEEE, 3–12.

[57] T. Kwon, B. Tekin, J. Stuhmer, F. Bogo, and M. Pollefeys. 2021. H2O: Two Hands Manipulating Objects for First Person Interaction Recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV).* IEEE Computer Society, Los Alamitos, CA, USA, 10118–10128. https://doi.org/10.1109/ICCV48922.2021.00998

[58] Taein Kwon, Bugra Tekin, Jan Stühmer, Federica Bogo, and Marc Pollefeys. 2021. H2o: Two hands manipulating objects for first person interaction recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision.* 10138–10148.

[59] Yann Labbé, Lucas Manuelli, Arsalan Mousavian, Stephen Tyree, Stan Birchfield, Jonathan Tremblay, Justin Carpentier, Mathieu Aubry, Dieter Fox, and Josef Sivic. 2022. MegaPose: 6D Pose Estimation of Novel Objects via Render & Compare. *arXiv preprint arXiv:2212.06870* (2022).

[60] Gun A Lee, Mark Billinghurst, and Gerard Jounghyun Kim. 2004. Occlusion based interaction methods for tangible augmented reality environments. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry.* 419–426.

[61] Gunnar Liestøl and Andrew Morrison. 2013. Views, alignment and incongruity in indirect augmented reality. In *2013 IEEE International Symposium on Mixed and Augmented Reality-Arts, Media, and Humanities (ISMAR-AMH).* IEEE, 23–28.

[62] Chuan-en Lin, Ta Ying Cheng, and Xiaojuan Ma. 2020. Architect: Building interactive virtual experiences from physical affordances by bringing human-in-the-loop. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems.* 1–13.

[63] Marcio C de F Macedo and Antonio L Apolinario. 2021. Occlusion handling in augmented reality: Past, present and future. *IEEE Transactions on Visualization and Computer Graphics* (2021).

[64] Maurizio Maisto, Claudio Pacchierotti, Francesco Chinello, Gionata Salvietti, Alessandro De Luca, and Domenico Prattichizzo. 2017. Evaluation of wearable haptic systems for the fingers in augmented reality applications. *IEEE*

transactions on haptics 10, 4 (2017), 511–522.

[65] James McCrae, Karan Singh, and Niloy J Mitra. 2011. Slices: a shape-proxy based on planar sections. *ACM Trans. Graph.* 30, 6 (2011), 168.

[66] Chris McDonald and Gerhard Roth. 2003. *Replacing a mouse with hand gesture in a plane-based augmented reality system.* Citeseer.

[67] Ryo Suzuki Mehrad Faridan, Bheesha Kumari. 2023. "ChameleonControl: Teleoperating Real Human Surrogates through Mixed Reality Gestural Guidance for Remote Hands-on Classrooms". https://ryosuzuki.org/chameleon-control/. Online; accessed April 5 2023, to be published.

[68] Victor Rodrigo Mercado, Thomas Howard, Hakim Si-Mohammed, Ferran Argelaguet, and Anatole Lécuyer. 2021. Alfred: the haptic butler on-demand tangibles for object manipulation in virtual reality using an ethd. In *2021 IEEE World Haptics Conference (WHC)*. IEEE, 373–378.

[69] Kyzyl Monteiro, Ritik Vatsal, Neil Chulpongsatorn, Aman Parnami, and Ryo Suzuki. 2023. Teachable Reality: Prototyping Tangible Augmented Reality with Everyday Objects by Leveraging Interactive Machine Teaching. *arXiv preprint arXiv:2302.11046* (2023).

[70] Oculus. 2020. Oculus Quest 2. Retrieved April 4, 2021, from https://www.oculus.com/quest-2/.

[71] Ohan Oda, Levi J Lister, Sean White, and Steven Feiner. 2010. Developing an augmented reality racing game. In *2nd International Conference on INtelligent TEchnologies for interactive enterTAINment*.

[72] JoonKyu Park, Yeonguk Oh, Gyeongsik Moon, Hongsuk Choi, and Kyoung Mu Lee. 2022. Handoccnet: Occlusion-robust 3d hand mesh estimation network. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 1496–1505.

[73] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. 2019. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 165–174.

[74] Siyou Pei, Alexander Chen, Jaewook Lee, and Yang Zhang. 2022. Hand Interfaces: Using Hands to Imitate Objects in AR/VR for Expressive Interactions. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*. 1–16.

[75] Ben Piper, Carlo Ratti, and Hiroshi Ishii. 2002. Illuminating clay: a 3-D tangible interface for landscape analysis. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 355–362.

[76] Ivan Poupyrev, Desney Tan, Mark Billinghurst, Hirokazu Kato, Holger Regenbrecht, and Nobuji Tetsutani. 2001. Tiles: A mixed reality authoring interface. In *(2001) INTERACT 2001 Conference on Human Computer Interaction*.

[77] Pornthep Preechayasomboon and Eric Rombokas. 2021. Haplets: Finger-worn wireless and low-encumbrance vibrotactile haptic feedback for virtual and augmented reality. *Frontiers in Virtual Reality* 2 (2021), 738613.

[78] Xun Qian, Fengming He, Xiyun Hu, Tianyi Wang, and Karthik Ramani. 2022. ARnnotate: An augmented reality interface for collecting custom dataset of 3D hand-object interaction pose estimation. In *The 35th Annual ACM Symposium on User Interface Software and Technology* (Bend OR USA). ACM, New York, NY, USA.

[79] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. 2016. You Only Look Once: Unified, Real-Time Object Detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, Los Alamitos, CA, USA, 779–788. https://doi.org/10.1109/CVPR.2016.91

[80] Yu Rong, Takaaki Shiratori, and Hanbyul Joo. 2021. Frankmocap: A monocular 3d whole-body pose estimation system via regression and integration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 1749–1759.

[81] Maria V Sanchez-Vives and Mel Slater. 2005. From presence to consciousness through virtual reality. *Nature Reviews Neuroscience* 6, 4 (2005), 332–339.

[82] Orit Shaer, Eva Hornecker, et al. 2010. Tangible user interfaces: past, present, and future directions. *Foundations and Trends® in Human–Computer Interaction* 3, 1–2 (2010), 4–137.

[83] Dandan Shan, Jiaqi Geng, Michelle Shu, and David F Fouhey. 2020. Understanding human hands in contact at internet scale. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9869–9878.

[84] Jinwook Shim, Yoonsik Yang, Nahyung Kang, Jonghoon Seo, and Tack-Don Han. 2016. Gesture-based interactive augmented reality content authoring system using HMD. *Virtual Reality* 20 (2016), 57–69.

[85] Mel Slater, Andrea Brogni, and Anthony Steed. 2003. Physiological responses to breaks in presence: A pilot study. In *Presence 2003: The 6th annual international workshop on presence*, Vol. 157. Citeseer.

[86] Stereolabs. 2021. Zed mini. Retrieved April 4, 2021 from https://www.stereolabs.com/zed-mini/

[87] Mengmeng Sun, Weiping He, Li Zhang, and Peng Wang. 2019. Smart haproxy: a novel vibrotactile feedback prototype combining passive and active haptic in AR interaction. In *2019 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*. IEEE, 42–46.

[88] Omid Taheri, Nima Ghorbani, Michael J Black, and Dimitrios Tzionas. 2020. GRAB: A dataset of whole-body human grasping of objects. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part IV 16*. Springer, 581–600.

[89] James Vallino and Christopher Brown. 1999. Haptics in augmented reality. In *Proceedings IEEE International Conference on Multimedia Computing and Systems*, Vol. 1. IEEE, 195–200.

[90] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Yuanzhi Cao, and Karthik Ramani. 2021. Gesturar: An authoring system for creating freehand interactive augmented reality applications. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 552–567.

[91] Tianyi Wang, Xun Qian, Fengming He, Xiyun Hu, Ke Huo, Yuanzhi Cao, and Karthik Ramani. 2020. CAPturAR: An augmented reality tool for authoring human-involved context-aware applications. In *Proceedings of the 33rd Annual ACM Symposium on User Interface Software and Technology*. 328–341.

[92] Peter Weir, Christian Sandor, Matt Swoboda, Thanh Nguyen, Ulrich Eck, Gerhard Reitmayr, and Arindam Dey. 2012. BurnAR: Feel the heat. In *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 331–332.

[93] Bowen Wen and Kostas Bekris. 2021. Bundletrack: 6d pose tracking for novel objects without instance or category-level 3d models. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 8067–8074.

[94] Andy Wu, Derek Reilly, Anthony Tang, and Ali Mazalek. 2010. Tangible navigation and object manipulation in virtual environments. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. 37–44.

[95] Yukang Yan, Chun Yu, Xiaojuan Ma, Xin Yi, Ke Sun, and Yuanchun Shi. 2018. Virtualgrasp: Leveraging experience of interacting with physical objects to facilitate digital object retrieval. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[96] Lixin Yang, Kailin Li, Xinyu Zhan, Fei Wu, Anran Xu, Liu Liu, and Cewu Lu. 2022. OakInk: A Large-scale Knowledge Repository for Understanding Hand-Object Interaction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 20953–20962.

[97] Li Zhang, Weiping He, Zhiwei Cao, Shuxia Wang, Huidong Bai, and Mark Billinghurst. 2022. Hapticproxy: Providing positional vibrotactile feedback on a physical proxy for virtual-real interaction in augmented reality. *International Journal of Human–Computer Interaction* (2022), 1–15.

[98] Qian Zhou, Sarah Sykes, Sidney Fels, and Kenrick Kin. 2020. Gripmarks: Using hand grips to transform in-hand objects into mixed reality input. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[99] Minglu Zhu, Zhongda Sun, Zixuan Zhang, Qiongfeng Shi, Tianyiyi He, Huicong Liu, Tao Chen, and Chengkuo Lee. 2020. Haptic-feedback smart glove as a creative human-machine interface (HMI) for virtual/augmented reality applications. *Science Advances* 6, 19 (2020), eaaz8693.

[100] Christian Zimmermann and Thomas Brox. 2017. Learning to estimate 3d hand pose from single rgb images. In *Proceedings of the IEEE international conference on computer vision*. 4903–4911.