Gripmarks: Using Hand Grips to Transform In-Hand Objects into Mixed Reality Input

Qian Zhou[†*]
qzhou@ece.ubc.ca

Sarah Sykes[†]
sarah.sykes@fb.com

Sidney Fels[*]
ssfels@ece.ubc.ca

Kenrick Kin[†]
kenrickkin@fb.com

[†] Facebook Reality Labs, Redmond, WA, USA   [*] University of British Columbia, BC, Canada
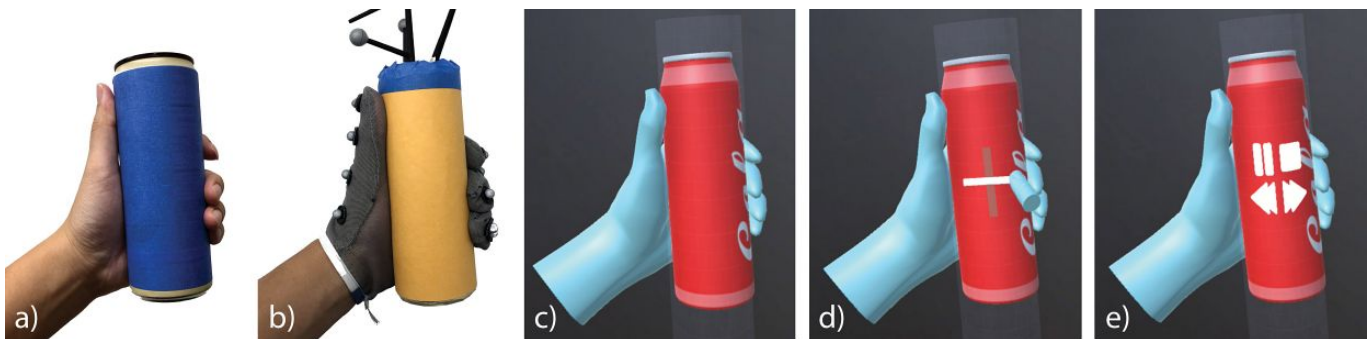
Figure 1. Gripmarks explore the concept of transforming an object a user is already holding, such as a soda can (a), into an input surface for mixed reality by recognizing the grip through hand tracking (b). The object is tracked only for visualization purposes and represents a real object in simulated augmented reality. The grip is recognized as a cylinder grip and the surface of the cylindrical object is estimated (c). The user employs an activation gesture on the estimated surface (d) to specify the size and location of a simple interface, such as a media player (e).

ABSTRACT

We introduce *Gripmarks*, a system that enables users to op-portunistically use objects they are already holding as input surfaces for mixed reality head-mounted displays (HMD). Leveraging handheld objects reduces the need for users to free up their hands or acquire a controller to interact with their HMD. Gripmarks associate a particular hand grip with the shape primitive of the physical object without the need of object recognition or instrumenting the object. From the grip pose and shape primitive we can infer the surface of the object. With an activation gesture, we can enable the object for use as input to the HMD. With five gripmarks we demonstrate a recognition rate of 94.2%; we show that our grip detection benefits from the physical constraints of holding an object. We explore two categories of input objects 1) tangible surfaces and 2) tangible tools and present two representative applications. We discuss the design and technical challenges for expanding the concept.

Author Keywords

gripmarks; grip recognition; tangible objects; mixed reality

CCS Concepts

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

INTRODUCTION

Mixed reality head-mounted displays (HMDs) allow users to view and interact with virtual content while mobile. Though hand and controller input are common ways for users to inter-act with their devices [17, 18, 20], users may not always have their hands free or controllers available. Throughout the day, users pick up, hold, and carry a variety of everyday objects such as food and beverage containers, bags, utensils, and other tools [7]. These objects prevent users from being able to use their hands to interact with their HMDs, yet they have the opportunity to provide inherent passive haptic feedback [15] for hand input.

Prior work has sought to opportunistically leverage existing objects in their environment as tangible props [13, 14, 28]. Although such work demonstrates the promise of using natural affordances, they often require object recognition and tracking or instrumenting the physical objects with sensors. In this work, we take a different perspective: through hand tracking, which is becoming popular in HMDs, we leverage information about the user's hand when gripping the object. We believe how the user grasps an object can reveal opportunities on how an uninstrumented object already in hand can be used [29].

To explore the feasibility and utility of enabling tangible inter-actions via hand grips, we developed *Gripmarks*, a system that enables users to opportunistically use objects they are *already* holding as mixed reality input, reducing the need to set aside

PREPRINT

their objects to free up their hands or acquire a controller. A *gripmark* associates or *bookmarks* a particular hand grip with interactions through understanding just the grip pose via hand tracking and without needing to instrument the object with sensors or performing additional object tracking (Figure 1).

From the grip pose we recognize and explore two categories of gripmark objects:

- *Tangible surface* - An object in one hand, typically the nondominant hand, is turned into a tangible surface for the other hand, typically the dominant hand, to interact with through touch.

- *Tangible tool* - An object in one hand, typically the dominant hand, is turned into a tap and swipe surface for the thumb or a trigger if the object is articulated.

To create a gripmark, we record hand pose data to create a new template and associate the template with a tangible surface primitive (cylinder, box, sphere) or tangible tool type (pistol, clicker).

When a gripmark is recognized, overlaid virtual content is rendered on the physical object to provide visual feedback. Users can perform an activation gesture to transform the object into an instant touchscreen or controller as input for the system or tied to a specific application of the HMD. To evaluate our approach as a proof of concept, we describe a system evaluation study with five physical objects. We found an average grip recognition rate of 94.2%, touch accuracy of 4.1 mm, and thumb swipe detection rate of 84.4%. Using the system, we present two simple applications of gripmarks with discussions on their usability. We conclude with a discussion on how to further extend the concept of gripmarks. We provide the following contributions:

1. We present a gripmark approach that supports interactions with everyday objects by using hand tracking to infer the object being held.

2. We detail proof-of-concept implementations to recognize different hand grips, detect touch input, and detect finger gestures such as pulling a trigger with the index finger and thumb swiping.

3. We evaluate the system's recognition and input accuracy across five physical objects and show results comparable to existing work. We show the importance of the physical object in the reliability of our grip recognition (94.2% recognition rate when holding the physical object compared to 39.0% when miming a grip).

4. We provide two illustrative applications of the interactions based on gripmarks and discuss their benefits and usability challenges.

Compared to approaches that detect the object surface through depth sensors on HMDs, our proposed approach has several strengths. First, to capture the object via the depth-camera on an HMD, the user must hold the object within the camera's field of view (FOV). This can be cumbersome for various interactions such as tapping on the object, as depth sensors often have small FOVs. Integration of depth sensors is also limited

by their power and cost. Our approach is instead as flexible as the hand tracking solution used. Second, for interactions that involve object manipulation, the camera may not capture the geometry of the in-hand object due to occlusion from the hand. Therefore, grip poses can help infer the geometry and associate it with possible interactions. Lastly, certain interactions can be performed eyes-free by the user, such as thumb swiping. With gripmarks, the user does not need to worry about where, relative to their body and HMD, they perform the interaction. For example, when using a tangible tool to advance slides in a presentation, users do not need to look at their hands nor keep their hands in the camera's FOV.

Gripmarks enables these types of interactions in new contexts. It takes a hand-centric view of turning already in-hand objects into interactable objects. Though we use hand tracking that requires markered gloves in this work, we believe our method and insights can be applied towards other hand tracking solutions, such as electromagnetic tracking of the hand pose with wearable technologies.

RELATED WORK
We discuss previous approaches to leveraging everyday objects as input for interactive systems and mixed reality.

Tangible User Interfaces in Mixed Reality
Tangible user interfaces (TUI) utilize the affordances of physical objects, surfaces, and spaces as means of interacting with digital information. As one of the pioneering works, Tangible Bits [15] proposed leveraging the affordances of everyday objects for seamless tangible interaction. Following Tangible Bits, researchers have repurposed everyday objects as tangible props [3, 6], created ephemeral user interfaces [28], and provided passive haptics using physical affordances already present in the domain environment [13].

Although TUIs provides natural passive haptics, it also requires users to pair the physical and virtual objects based on the application. Sometimes it would be impractical to find the physical object identical to its virtual counterpart, resulting in an inevitable mismatch between the virtual and real objects. In particular, this mismatch affects users' level of immersion as well as the believability of the experience in mixed realities [23]. Substitutional Reality investigated the mismatch between the virtual and physical proxies with different degrees of discrepancy. They found that objects with similar affordances in the manipulable parts are the best candidates for substitution [23]. Annexing Reality proposed the opportunistic use of everyday objects by finding the optimal mapping between the physical-virtual pairs based on their shapes [14]. Sparse Haptic Proxy investigated a generic approach of providing passive haptics by redirecting users' hands to a matching primitive of the proxy [4]. Other work investigated retargeting the passive haptics to support interactions with multiple virtual objects without precisely-located haptic proxy objects [1, 16].

Instead of categorizing and recognizing the physical objects to be held, we take a different perspective by focusing on the user's hands and how they grasp the object. Prior research has

also investigated the feasibility of transforming physical surfaces into touch surfaces by sensing hands [12, 22, 31]. They focus on touch interactions with tangible surfaces detected through depth and capacitive sensing, while we rely only on the tracking of the hands to enable interactions with handheld physical surfaces and tools.

Grip Sensing and Interactions

Understanding how humans utilize hands to grasp objects has long been a topic of interest, including examining the grip taxonomy [33], as well as the usage frequency of different grips in the daily life [27]. Depending on the context, a grip can convey meaningful information of user's intention based on the ways people hold the objects [29]. Researchers have explored using the information of hand grip to support interactions with the mobile devices [9, 32]. Graspables [26] explored the idea of using grips as input gestures. They developed graspable prototypes with the capability of sensing the hand grip as input. HandSense [30] employs capacitive sensors for detecting different grips with a handheld mobile display to support interactions such as touching and squeezing. MTPen [25] proposed to use how users grip the pen as a mechanism for model switching by detecting touch data on the pen. They combined different grips an gestures to enable new interaction techniques with the MTPen. Imaginary Interfaces [10] uses an 'L' hand pose to create reference points for in-air interactions. By leveraging hand pose to initiate interactions, the approach does not require visual feedback. Other studies have investigated using grips to support active reading [32], predict touch point [19], and adapt UIs [9] on mobile devices such as tablets and smartphones.

Though our work also uses grips for interactions, we do not focus on individual mobile devices with embedded sensors to detect grips. Instead, we opportunistically turn everyday handheld objects into instant user interfaces by using the grip information from hand tracking.

Hand Tracking

The importance of hand tracking has been long appreciated in the practice of Human Computer Interactions. There has been a tremendous body of literature related to tracking technologies, gesture taxonomies, and recognition. Comprehensive surveys on hand tracking and gesture recognition can be found in [2, 5, 21]. Sensor-based hand tracking requires use of sensors such as inertial measurement unit or ultrasonic sensors mounted on the hands to capture the movement of hands and fingers, while vision-based hand tracking requires the acquisition of the hand images using cameras [5]. Both approaches have their advantages and disadvantages. Sensor-based tracking can be uncomfortable to users as sensors are mounted on hands with physical contact, while vision-based tracking is more user friendly but suffers from configuration complexity and occlusion problems [21]. Recent advances of optical marker-based motion capture provides high-fidelity hand tracking with frequent occlusions [11], which we use in our work to detect grips. Although we use an optical motion capture system, our approach is agnostic to the precise hand-tracking method and could also use sensor-based tracking for grip detection.
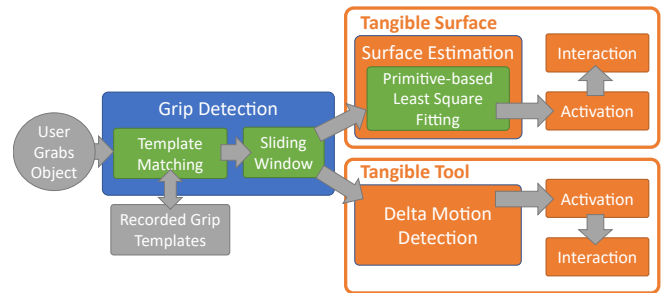

Figure 2. Gripmark detection and activation flow.

GRIPMARKS

A gripmark is constructed through grip template generation and specification of the object primitive. When a user grabs an object we match the grip with saved gripmarks. Once recognized, the user can enable the object through an activation gesture. A summary of the process is shown in Figure 2.

For tangible surfaces, we demonstrate three types of surfaces with three primitives: cylinder, box, and sphere. Bottles, cans and mugs are often cylindrical. Books, wallets, electronic devices are boxy. Balls and round fruit are spherical. Our representative objects for each category are a soda can, a clutch wallet, and an apple for cylinder, box, and sphere respectively. For tangible tools, we demonstrate two tool grips: pistol and clicker. Pistol grips are operated using index (and middle) finger trigger pulls while clicker grips are operated with thumb taps/swipes. These grips and objects are shown in Figure 3.

For each gripmark we create, first we specify the grip type and then record pose sequences of the user holding the object and compute a template. A template is the collection of each joint angle's mean plus a 95% confidence interval from 200 frames of poses.

A gripmark is detected through template matching if the normalized correlation coefficient exceeds a threshold of 0.95. As shown in Figure 2, we use a sliding window of 120 frames (1.33 s at 90 Hz) of pose data and if at least 80% of the frames are of the same gripmark with the highest likelihood, then we consider that gripmark to be detected.

Once we detect the gripmark, each gripmark category, tangible surface and tangible tool, has a different activation mechanism and interaction technique. The activation mechanism serves two purposes. The first is to allow the user to indicate their intent to use the object and reduce the Midas Touch problem of accidentally triggering input events. The second is to allow the user to specify and orient the interactable region.

When the user releases the object, the current gripmark interaction is ended. We detect a release when the sum of the joint angles is less than 60% of the sum of the joint angles in the template. A subsequent gripmark can only activate after the previous gripmark has been released so small shifts in the hand grip cannot switch to a similar, but different gripmark. This also matches the expectation that the user would have to let go of the current object in order to grab a different object.

Figure 3. Tangible surfaces have three grip types: a) cylinder, b) box, and c) sphere. The estimated surfaces are shown in wireframes. Tangible tools have two grip types: d) pistol and e) clicker.
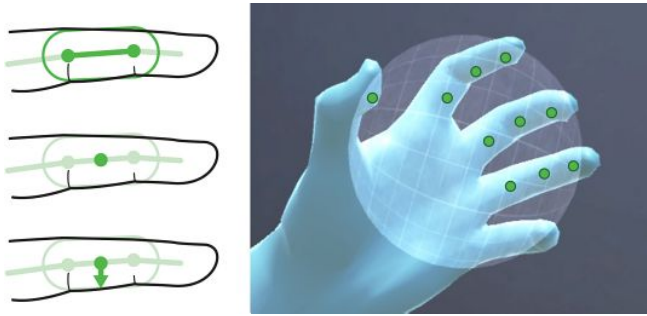

Figure 4. Contact points are generated by projecting the center of the finger bones towards the palm side of the fingers. The object surface is then estimated from the contact points.

Figure 5. The user activates the surface by drawing a '+' that also sets the size, position, and orientation of the active region.

Tangible surface

Knowing the primitive associated with the pose, we compute a set of contact points from the hand to estimate the surface of the object. For each primitive, we use a different set of contact points. The contact points are estimated by projecting from the middle of the finger bones to the palm side surface of the fingers (Figure 4).

The bones we use to create contact points are:

- cylinder - thumb {metacarpal, proximal}; index {proximal, intermediate, distal}; middle {proximal, intermediate, distal}; ring {proximal, intermediate, distal}; pinky {intermediate}

- box - index {intermediate, distal}; middle {intermediate, distal}; ring {intermediate, distal}; pinky {intermediate, distal}

- sphere - thumb {distal}; index {proximal, intermediate, distal}; middle {proximal, intermediate, distal}; ring {proximal, intermediate, distal}

Using the estimated contact positions, we fit the object. For the box and sphere we fit the surface using a linear least squares method. We fit the cylinder with a nonlinear least squares method using the Levenberg-Marquardt algorithm with the initial guess based on last frame. It converges within 20 steps for each frame.

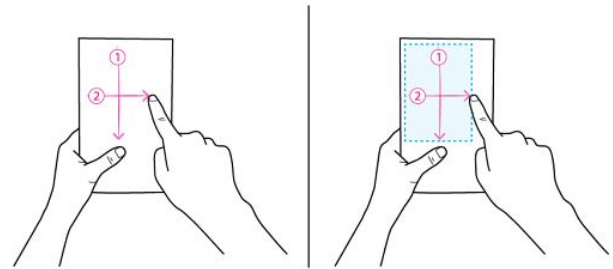Note for the cylinder grip, we estimate a taller than necessary cylinder, with the assumption that the user will interact with the physical portion, i.e., what they can see, of the object. Relatedly, we use the finger contact points to estimate the plane of the box but position it at the front side using the thumb distal joint position. We slightly inflate the estimated surface by 10% so that the interacting finger can reach the estimated surface, and not be blocked by the physical object. We visualize the estimated surface as wireframes in Figure 9.

Activation

After we estimate the surface of the object, the user can activate the object by drawing a '+' on the estimated surface (Figure 5). The '+' determines the size, orientation, and location of the activated region of the object. We can also align the activated region with the natural up direction of the object, such as the axis of the cylinder. Since the estimated surface is slightly inflated to ensure contact detection, we can further use the fingertip positions to refine the position or inflation of the surface.

Tangible tool

A gripmark tangible tool currently has two possible types: pistol and clicker. When the pose is recognized we use a sliding window to detect a static pose, where the finger local positions have standard deviations smaller than a threshold of 2 mm with respect to the hand. When a static pose is established, we look for delta motions based on the gripmark type. For the pistol gripmark, we look for index or middle finger motions where the distal joint position moves towards the wrist relative from rest. For the clicker grip, we look for the thumb tip lifting away from the hand.
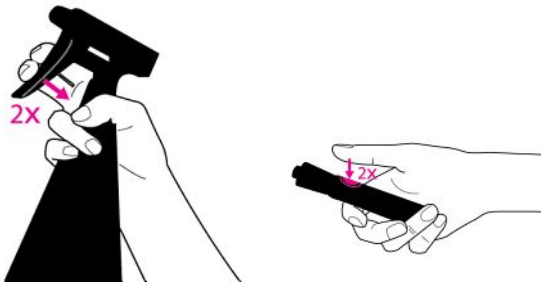
PREPRINT

Figure 6. The user activates the tool by performing a double delta motion. The delta motion also sets the rest position for gesture detection.

Figure 7. Left: The five objects are visualized through motion capture tracking. Right: The hands are tracked using markered gloves and the index finger is tracked with a single rigid body.

Activation

Activating the tool requires a double-press (of the trigger) or a double-tap (of the surface) as shown in Figure 6. From the double-press we can determine the size of the trigger pull and from the double-tap we can determine the rest position of the thumb. The size of the pull can then be used as a threshold for determining the point of future trigger pulls and the thumb rest position can be used to determine where the swipes and taps begin.

SYSTEM EVALUATION

We evaluated our template matching approach and input recognition. *Gripmarks* is implemented using the Unity game engine running on a Falcon Northwest PC with an Intel Core i7-5930K CPU and NVIDIA GeForce GTX TITAN X. We use an Oculus Rift and motion-capture-based hand tracking system [11] that requires users to wear gloves with retro-reflective markers tracked using 18 Optitrack cameras. We also use motion capture to track our real physical objects so that we can render them in VR (Figure 7). However, it is important to note that we use motion capture only to visualize the physical objects as one might see them in the real world. Our interactions are built upon hand tracking information and we visualize simulated AR interfaces by rendering them additively over the VR environment.
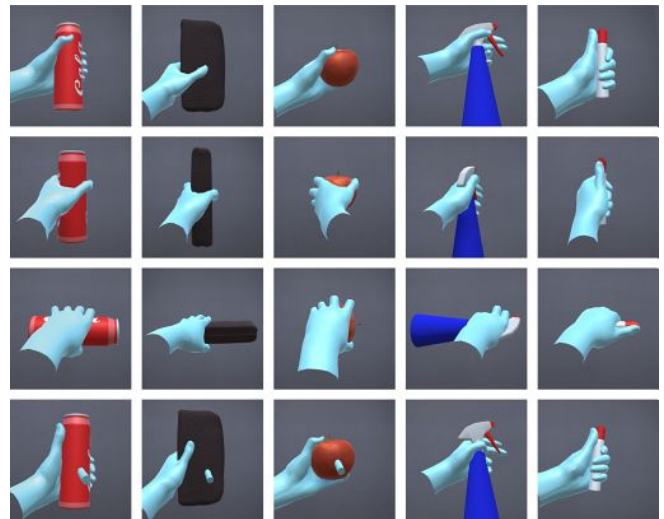

Figure 8. For each of the five grip types (columns), we record four templates (rows).

Due to variations in hand size and the way individuals hold objects, our system requires that users have personalized templates. We recruited eight participants from within Facebook Reality Labs (four male and four female) and had them record grip templates for each object and then had the system recognize their grips.

For each participant we recorded templates for five grips (cylinder, box, sphere, pistol and clicker) corresponding to five physical objects (soda can, clutch wallet, apple, spray bottle and dry erase marker) as shown in Figure 3. The participant held the objects in four different poses and we created a template for each object pose. We showed the participant pictures of grips holding each object (Figure 8) and they performed each grab sequentially based on pictures. We instructed the participant to adjust their grips slightly such as tightening the grips or imagining touching the surface with the other hand to help capture the grip variations when they interacted with the objects. Thus, we recorded twenty total templates per participant, four for each of the five grips and corresponding objects.

Grip Recognition

After template recording, the participant picked up each of the five objects to test our grip recognition. For each object, the participant picked up the object, held it for three seconds, and set it back down. This process was done ten times per object, resulting in 30 seconds of grip recognition per object. After going through all five objects, the participant then *pretended* to hold the object by forming the grip pose without actually grabbing any physical object for three seconds per object.

With the five grips, our system has a recognition rate of 94.2% while holding the physical object (Table 1). If the user only pretends to hold the object the recognition rate is 39.0% (Table 2). This suggests that holding the physical object reinforces the reliability of recognizing gripmarks. The physical object serves as a physical guide for the fingers to recreate the saved pose and also helps keeps the fingers locked in place for stable

PREPRINT

	None	Cyl.	Plane	Sphere	Pistol	Click.
Cylinder	3.9	**94.8**	1.3	0.0	0.0	0.0
Plane	4.4	0.8	**92.7**	2.1	0.0	0.0
Sphere	0.2	2.2	0.0	**97.6**	0.0	0.0
Pistol	2.8	7.4	0.0	0.0	**89.2**	0.6
Clicker	0.9	1.0	1.2	0.0	0.0	**96.8**

Table 1. Recognition rates for grips with physical object in hand.

	None	Cyl.	Plane	Sphere	Pistol	Click.
Cylinder	26.6	**33.7**	0.0	13.3	26.5	0.0
Plane	62.7	12.0	**25.3**	0.0	0.0	0.0
Sphere	33.7	12.3	0.0	**41.4**	12.6	0.0
Pistol	29.1	13.9	0.0	0.0	**37.9**	19.1
Clicker	20.2	9.0	0.0	0.0	13.9	**56.9**

Table 2. Recognition rates for grips without physical object in hand.

recognition. The recognition rate of 94.2% compares well to past grip recognition systems such as 95% for Graspables [26] and 87% for MTPen [25].

The results of Table 1 and Table 2 demonstrate the integral role of a physical object in both its opportunistic use as an interaction surface and its positive effect on grip recognition. By having the physical object in the loop, the user can more consistently make the same grip, which allows more reliable recognition of similar but different grips. Without the object, the user's hand has nothing solid to conform to and thus the grips become less discernible. But, gripmarks is not mutually exclusive to midair techniques as midair interfaces can be utilized when the user's hands are free similar to Imaginary Interfaces [10]. The Midas Touch problem is also a general concern of gestural interfaces. With physical objects, the interactions can be constrained to the physical object which the user can tactilely feel. When interacting in midair without the object, the interactive bounds are less clear, and the user needs to be more mindful of their intentions.

Tangible Surface Tap Detection

We tested the tap recognition of our estimated surfaces under two conditions: targets rendered on 1) the grip estimated surface and 2) the motion capture surface. We used the object surface tracked by motion capture as the baseline for tap detection. The targets were nine squares with 1.67 cm sides arranged in a 3x3 grid.

For each condition, the participant grabbed the object and tapped the highlighted target on the grid with the index finger on their dominant hand (Figure 9). The targets were presented in randomized order, with each target position appearing twice for a total of 18 taps. The participant performed these taps on all three tangible surface objects. We counterbalanced the ordering of the two target render conditions.

We measured tapping accuracy as the distance between the target position and tapped position. The tapped position was where the finger contacted the estimated surface and the motion capture surface respectively for the two conditions. The target distance error for the grip estimated surface is 4.1 mm
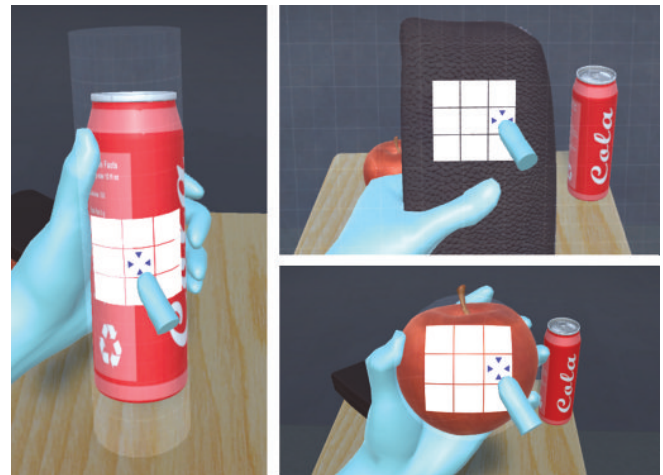

Figure 9. Surface estimation of each primitive type is shown in wireframe. Note that the opaque objects are shown to visualize the real objects (tracked with motion capture), but the surface estimation does not utilize information about the motion capture tracked physical objects.

and 3.1 mm for the motion capture surface. We found the detection rate for detecting taps on the grip estimated surface to be 91.7%. We kept false positives for taps on the estimated surface to just 0.30% by adding hysteresis, i.e., the participant had to lift their finger by 0.625 cm from the estimated surface to begin a new press. The finger lift threshold likely caused missed presses if the participant did not lift their finger enough before the next press. The result is comparable to prior work, e.g., MRTouch [31], which detects tapping on physical surfaces with spatial accuracy of 5.4 mm and 95% capture rate for 1.6 cm buttons.

Tangible Tool Delta Detection

We evaluated delta detection by having the participant grab the object (spray bottle or dry erase marker), hold it, and perform the delta gestures (double-pull for spray bottle and double-tap on the marker in Figure 10) four consecutive times before setting the object down before the next repetition. The participant performed this procedure five times per object, thus performing the action 20 times per object. We found a detection rate of 81.9% with no false positives. The participant also performed the same procedure for each left and right thumb swipe, for an 84.4% detection rate and a 7.5% false positive rate. The 81.9% and 84.4% detection rates are lower than the detection rates of 93.2% for double-taps and 94.7% for swipes on MTPen [25] respectively. A source of false positives occurred when the participant returned the thumb back to a neutral position; overshooting the neutral position would sometimes trigger a false swipe in the return direction.

EXAMPLE APPLICATIONS

To better understand the potential and limitations of interactions derived from gripmarks we implemented two applications, one using a tangible surface and one using a tangible tool. We turned a tangible surface into a calculator operated by pressing buttons on the surface of a clutch wallet (Figure 11 (Left)). We used the thumb swiping surface of a tangible tool
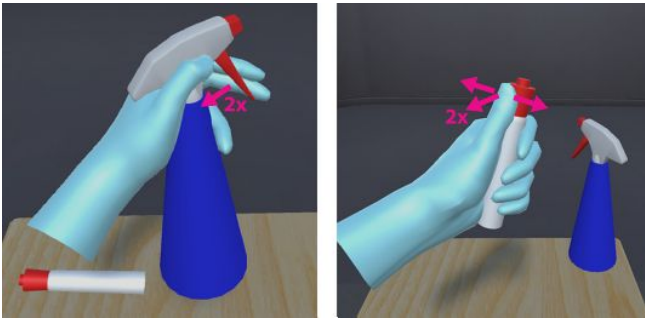
Figure 10. Delta motion on tangible tools. Left: Double-pull of the spray bottle trigger. Right: Swipe and double-tap on the dry erase marker.
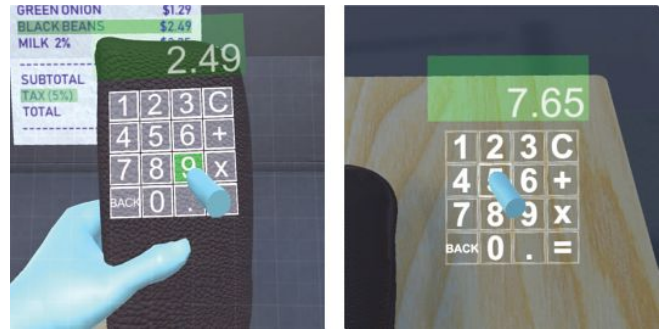

Figure 11. Calculator buttons are placed on the estimated plane of the clutch wallet (Left) and in midair anchored to the world (Right). The dominant index fingertip is tracked using a single motion capture rigid body as shown in Figure 7 (Right).
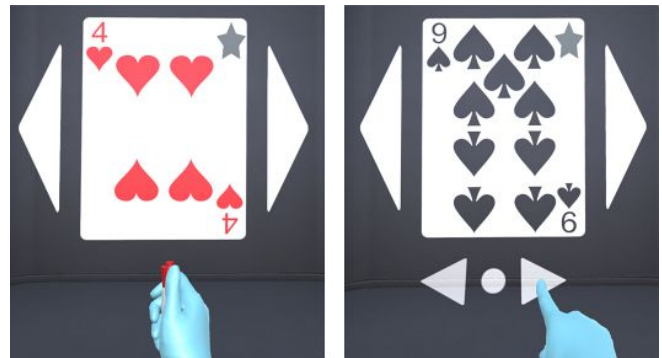

Figure 12. Navigate left and right and select are performed by thumb swiping and double tapping on the dry erase marker (Left) and by pushing on midair buttons anchored in the world (Right).

to perform left and right directional navigation and double-tap for selection in a card matching task (Figure 12 (Left)); the interactions are representative of navigating a photo browser or slide presentation. We recruited 18 participants external to our organization (ten male and eight female) to try these inputs, as well as corresponding midair button interfaces (Figure 11 and Figure 12 (Right)) as a straightforward point of comparison.

Calculator on tangible surface

We chose a clutch wallet as our surface; it has a large enough surface to comfortably fit a calculator with 16 buttons of 2 cm width. The surface is slightly challenging as the surface is uneven and the front surface isn't perfectly parallel with the back surface estimated from the finger contact points. The uneven nature of the clutch wallet emphasizes the benefit of using the tapping hand to perform an activation gesture. The deepest points of the activation reveal the true physical points on the surface.

We had participants start from the box grip and asked them to make three initial taps with the index finger on their dominant hand as a simplified proxy of the activation gesture. From the three points we estimated the location of the physical surface, which provided a better estimate of the plane orientation for calculator placement. As our hand tracking does not precisely measure the thickness of a user's hand, we erred on pushing the plane away from the object and towards the user to reduce the chance of it getting "stuck" inside the physical object. This led to an average of 1.02 cm distance between the estimated surface and the physical surface.[1]

To give participants a clear task to perform, we instructed them to add up three items from a grocery receipt and apply a tax to the total. They were required to get the final number correct. Participants performed the task for about 10 minutes.

While not a formal study, we summarize the general observations about participant experience with the calculator task: 1) Most participants felt the tactile feedback from feeling the wallet increased their satisfaction. Five participants specifically mentioned the tactile feedback is a good experience with the finger rebounding off the wallet. 2) We observed some

challenges with physical comfort of holding the object over prolonged periods. The weight of the object impacts comfort: heavier objects can become uncomfortable to hold for long periods. Three participants mentioned physical stress when holding the wallet. 3) The estimation of the surface holds if the participant keeps their grip consistent; we found users adjusting their grip after holding the wallet for too long which would cause the plane to shift. Depending on the weight and size of the object, using the tangible surface might be better for shorter periods of time with reactivation being necessary at the start of each chunk of interactions.

We recorded the completion time to perform the calculator task and computed the number of taps per minute.[2] We also had users perform the task for 10 mins using a midair calculator anchored to the world. As with the wallet, participants used their index finger on the dominant hand to press the buttons on the calculator (Figure 11). We found the performance to be comparable between the input types: 67.5 taps/min and 70.7 taps/min for the midair and wallet calculator (Figure 13 (Left)). The error rate (hitting the wrong button) was 0.9% and 1.3% for the midair and wallet respectively. Our approach is usable while providing the sensation of passive haptic feedback from

---

[1]To help participants make faster sequential button presses, we lowered the unpress hysteresis threshold to 0.375 cm from the system evaluation's 0.625 cm.

[2]We removed 6.14% tap outliers using Generalized Extreme Studentized Deviate test when the interval between two taps is too large.
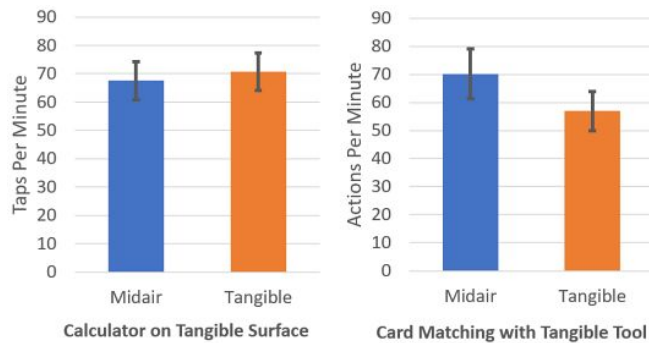
PREPRINT

Figure 13. Participants' performance when (Left) using a calculator on tangible surface and (Right) card matching with tangible tool compared to midair interactions with means and 95% confidence intervals.
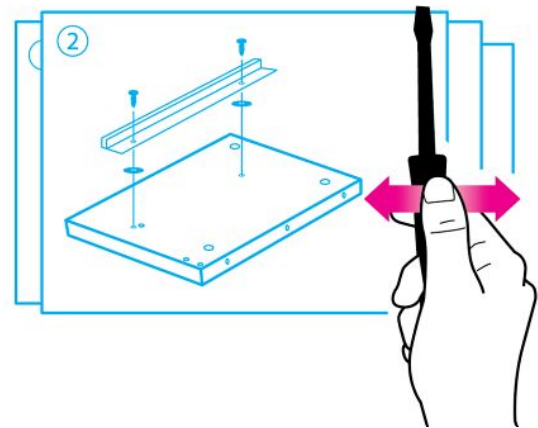

Figure 14. If the system recognizes the tool as a screwdriver and the piece of furniture being assembled, the system can automatically bind the tangible tool input for navigating assembly instructions.

an object the user is already holding. There is still room for refinement with better contact detection to provide tighter coupling between button press and physical contact, which can lead to performance gains over midair as Dudley et al. found for text entry [8].

Though participants liked not having to hold anything to use midair buttons, one of the key advantages of having interfaces on the physical object already being held is the user always knows where the input surface is. Midair buttons, however, need to be intelligently placed by the system, otherwise the user may have a difficult time locating them or using them comfortably. Context switching between a midair UI and other visual content means users need to reacquire the midair buttons frequently. Also, midair buttons take up additional visual real estate, which can potentially occlude regions of interest.

Card matching with tangible tool

We explored the use of a dry erase marker as a tangible tool using the clicker grip. We used similar swipe and tap interactions as described in MTPen [25], but our detection relies on hand tracking instead of sensors embedded on the pen. The card matching task required participants to cycle through eight pairs of cards and find each matching pair. During the task, participants performed left and right thumb swipes to navigate left and right and double tap to select the card to match. They performed the task for about 10 mins.

For swipe and double tapping interactions we observed the following challenges: 1) Based on their thumb mobility, some participants had a harder time swiping their thumb towards the direction of the index finger's knuckle (vs the tip). Four participants mentioned the finger movements were awkward and unnatural. 2) Our detection is based on starting swipes from the initial rest pose. However, some participants would overshoot the rest pose and cause a false positive in the other direction. 3) For double tapping, the inclusion of the marker appeared to affect the thumb's mobility of the thumb tap. With a thick object, the thumb has less room to make delta motions, as the object occupies the space that the thumb can lift above the side of the index finger. 4) Like the tangible surface wallet, adjustments in grip can affect the reference points for starting a tap or a swipe. Similar to the findings from MTPen [25], the

second tap of double tapping gesture sometimes shifted from the first tap, causing failure of detection.

An advantage of thumb swipes is users can perform them eyes free, as commented by a participant, "I did not have to look at the floating control to see where I was pressing." The participants did not need to look at their hands and additional UI elements in the visual display, such as midair buttons, are not needed. The eyes-free interaction allows users to focus on the primary task without dividing attention. As the swipes are performed with finger motions relative to the hand, it seems possible for users to also perform the task while on the move.

As with the calculator task, we measured the completion time and computed the number of actions (swipe or double tapping) per minute.[3] We also had users perform the task for 10 minutes with a midair buttons anchored in the world. We found that the midair input (70.2 actions/min) was more efficient than the marker input (56.9 actions/min) shown in Figure 13 (Right). Despite the technical challenges, using our tangible tool was within 81.1% of a simple midair solution. Ultimately the advantage of the tangible tool is that the user can perform eyes-free actions on the tool without needing to free their hand, for example, the user could navigate a furniture assembly tutorial while holding a screwdriver for turning screws (Figure 14). Also, unlike midair buttons, the tangible tool interface does not take up any additional visual real estate.

DESIGN EXTENSIONS

To increase the practicality of *Gripmarks* we discuss design opportunities: working around the hand to display information, binding multiple applications to a gripmark, extending the system with contextual awareness of the specific object and situation, and leveraging more physical affordances.

Currently, we simply render the estimated surface whenever a gripmark is detected. In practice, a user may just like to use the object for their original intention, such as drinking from a soda can, without the visual distraction. The system should

---

[3]We removed 6.06% action outliers using Generalized Extreme Studentized Deviate test when the interval between actions is too large.

subtly signal to the user when a gripmark is available, such as revealing an icon of the bound application in the open region of the hand between the thumb and fingertips. After activating the surface, if the open region is too small for the intended interface on the tangible surface, we can extend the menu by placing midair UI on the outside of the hand (Figure 15 Left). The interplay between interactions on surface and midair can be further explored.

Though our current implementation hardcodes the mapping between a gripmark to an application (e.g., box grip to calculator) users could optionally bind and choose from multiple applications corresponding to a gripmark. This will be necessary if the user often holds similarly shaped objects where a gripmark cannot distinguish between different cans of similar circumference. Adding a simple application picker (i.e., buttons) to the object surface is a simple way for the user to choose from multiple application. Alternatively, different grips or activation gestures on the same object could determine the application or context. For example, in addition to the clicker grip, a tripod grip can be associated with the marker and an index finger delta motion can activate the marker to be a virtual pointer or 3D stylus. The marker can still be used as a thumb swipe surface when held with the clicker grip.

More interestingly is how to include contextual information to the system. The motivation of gripmarks is centered around needing to only understand the hand and inferring the object. However, if simple object character recognition can identify the particular object (e.g., reading the label on an object) we can add context-aware menus to the object. For example a gripmark for holding a book can reveal the user's interactive notes for the particular book if the book title is recognized. The object recognition can be lightweight as it only needs to identify the object and not track its full six positional/rotational degrees of freedom. Recognizing that a user is holding a slot head screwdriver instead of a Phillips head (Figure 14), the system could highlight the steps that require the slot head or could suggest to the user to change tools.

Lastly gripmarks can leverage more of the objects inherent physical affordances. We presented the spray bottle trigger, which can be used as a raycast selector (Figure 15 (Right)). Other articulated components on common objects, such as the push button on a retractable pen, can be mapped to a discrete action when operated by the thumb in a way similar to [6]. Physical objects can also be rotated to serve as a virtual dial.

TECHNICAL IMPROVEMENTS AND FUTURE WORK

Gripmarks is comprised of several components with interesting technical challenges. Our current system requires users to wear gloves and a motion capture system. More practical alternatives to tracking the hand while holding objects include using gloves with sensors or wearables with EMG detection [5]. This information can be combined with inside-out optical tracking of hands.

As our method currently requires a tightly coupled grip, squishy objects could lead to shaky grips and be difficult to template match. Also, oddly shaped objects that do not lend themselves to be grabbed in a consistent way would be poor
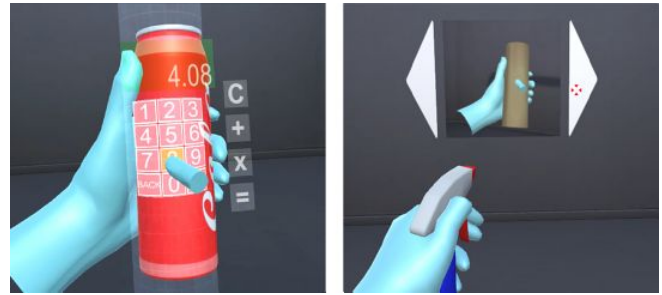

Figure 15. Left: On objects with smaller surface area, buttons can be split between the surface and midair. Knowing where the hand is, the midair overflow buttons can be placed to the side of the hand. Right: The spray bottle can be used as a virtual pointing, using the hand orientation to point and the trigger gesture to select.

candidate objects for gripmarks. Grip detection could be further improved through advanced machine learning techniques, which could expand the types of usable objects. Better grip detection could also lead to more similar but distinguishable gripmarks, increasing the specificity of interactions. Tangible surface touch interaction can improve by better estimating the time of contact with the surface, so taps (and drags) occur at the time of physical contact. Thumb taps and swipes on tangible tools can be improved with more sophisticated heuristics or gesture recognition through deep learning [24]. And, as mentioned previously, even simple additions of context awareness of the object details and user intention can improve the convenience of gripmarks.

We compared midair buttons to gripmarks to show the contrast between using an interface when the hands are not holding anything versus when they are already holding an object and thus can opportunistically use the object instead of having to set it down. As future work, it would be interesting to see how gripmarks compare to interacting with a mimed grip (i.e., without object) or with real devices such as a physical calculator.

CONCLUSION

We have developed and presented *Gripmarks* an end-to-end system for opportunistically turning everyday objects users are already holding into mixed reality input. Our approach uses hand tracking to associate hand grips to different categories of handheld objects without the need to instrument or further track the objects. Gripmark objects can then be used as tangible surfaces or tangible tools. We have presented proof-of-concept implementations of the components required to realize gripmarks. Our work highlights technical paths for improving the fidelity of the overall system. Future work can also build upon gripmarks and increase their utility by expanding the objects and contexts in which gripmarks can be used.

ACKNOWLEDGMENTS

REFERENCES

[1] Mahdi Azmandian, Mark Hancock, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2016. Haptic retargeting: Dynamic repurposing of passive haptics for enhanced virtual reality experiences. In *Proceedings of the 2016 chi conference on human factors in computing systems*. ACM, 1968–1979. DOI: `http://dx.doi.org/10.1145/2858036.2858226`

[2] Hong Cheng, Lu Yang, and Zicheng Liu. 2015. Survey on 3D hand gesture recognition. *IEEE transactions on circuits and systems for video technology* 26, 9 (2015), 1659–1673. DOI: `http://dx.doi.org/10.1109/TCSVT.2015.2469551`

[3] Kai-Yin Cheng, Rong-Hao Liang, Bing-Yu Chen, Rung-Huei Laing, and Sy-Yen Kuo. 2010. iCon: utilizing everyday objects as additional, auxiliary and instant tabletop controllers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1155–1164. DOI: `http://dx.doi.org/10.1145/1753326.1753499`

[4] 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*. ACM, 3718–3728. DOI: `http://dx.doi.org/10.1145/3025453.3025753`

[5] Ming Jin Cheok, Zaid Omar, and Mohamed Hisham Jaward. 2019. A review of hand gesture and sign language recognition techniques. *International Journal of Machine Learning and Cybernetics* 10, 1 (2019), 131–153. DOI: `http://dx.doi.org/10.1007/s13042-017-0705-5`

[6] Christian Corsten, Ignacio Avellino, Max Möllers, and Jan Borchers. 2013. Instant user interfaces: repurposing everyday objects as input devices. In *Proceedings of the 2013 ACM international conference on Interactive tabletops and surfaces*. ACM, 71–80. DOI: `http://dx.doi.org/10.1145/2512349.2512799`

[7] Andy Crabtree and Peter Tolmie. 2016. A Day in the Life of Things in the Home. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. ACM, 1738–1750. DOI: `http://dx.doi.org/10.1145/2818048.2819954`

[8] John Dudley, Hrvoje Benko, Daniel Wigdor, and Per Ola Kristensson. 2019. Performance envelopes of virtual keyboard text input strategies in virtual reality. In *Proceedings of the 17th IEEE International Symposium on Mixed and Augmented Reality*. 289–300. DOI: `http://dx.doi.org/10.1109/ISMAR.2019.00027`

[9] Rachel Eardley, Anne Roudaut, Steve Gill, and Stephen J Thompson. 2018. Designing for Multiple Hand Grips and Body Postures within the UX of a moving Smartphone. In *Proceedings of the 2018 Designing Interactive Systems Conference*. ACM, 611–621. DOI: `http://dx.doi.org/10.1145/3196709.3196711`

[10] Sean Gustafson, Daniel Bierwirth, and Patrick Baudisch. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In *Proceedings of the 23nd annual ACM symposium on User interface software and technology*. ACM, 3–12. DOI: `http://dx.doi.org/10.1145/1866029.1866033`

[11] Shangchen Han, Beibei Liu, Robert Wang, Yuting Ye, Christopher D Twigg, and Kenrick Kin. 2018. Online optical marker-based hand tracking with deep labels. *ACM Transactions on Graphics (TOG)* 37, 4 (2018), 166. DOI: `http://dx.doi.org/10.1145/3197517.3201399`

[12] Chris Harrison, Hrvoje Benko, and Andrew D. Wilson. 2011. OmniTouch: Wearable Multitouch Interaction Everywhere. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology*. ACM, 441–450. DOI: `http://dx.doi.org/10.1145/2047196.2047255`

[13] 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*. ACM, 211–218. DOI: `http://dx.doi.org/10.1145/1450579.1450625`

[14] Anuruddha Hettiarachchi and Daniel Wigdor. 2016. Annexing reality: Enabling opportunistic use of everyday objects as tangible proxies in augmented reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 1957–1967. DOI: `http://dx.doi.org/10.1145/2858036.2858134`

[15] 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*. ACM, 234–241. DOI: `http://dx.doi.org/10.1145/258549.258715`

[16] Luv Kohli, Mary C Whitton, and Frederick P Brooks. 2013. Redirected Touching: Training and adaptation in warped virtual spaces. In *2013 IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE, 79–86. DOI: `http://dx.doi.org/10.1109/3DUI.2013.6550201`

[17] Magic Leap. 2019. Magic Leap One: Creator Edition | Magic Leap. (2019). `https://www.magicleap.com/magic-leap-one`.

[18] Microsoft. 2019. Microsoft HoloLens | Mixed Reality Technology for Business. (2019). `https://www.microsoft.com/en-us/hololens`.

[19] Matei Negulescu and Joanna McGrenere. 2015. Grip change as an information side channel for mobile touch interaction. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 1519–1522. DOI: `http://dx.doi.org/10.1145/2702123.2702185`

[20] Oculus. 2019. Oculus. (2019). `https://www.oculus.com`.

[21] Siddharth S Rautaray and Anupam Agrawal. 2015. Vision based hand gesture recognition for human computer interaction: a survey. *Artificial intelligence review* 43, 1 (2015), 1–54. DOI: `http://dx.doi.org/10.1007/s10462-012-9356-9`

[22] Munehiko Sato, Ivan Poupyrev, and Chris Harrison. 2012. Touché: enhancing touch interaction on humans, screens, liquids, and everyday objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 483–492. DOI: `http://dx.doi.org/10.1145/2207676.2207743`

[23] Adalberto L Simeone, Eduardo Velloso, and Hans Gellersen. 2015. Substitutional reality: Using the physical environment to design virtual reality experiences. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3307–3316. DOI: `http://dx.doi.org/10.1145/2702123.2702389`

[24] Mohamed Soliman, Franziska Mueller, Lena Hegemann, Joan Sol Roo, Christian Theobalt, and Jürgen Steimle. 2018. FingerInput: Capturing Expressive Single-Hand Thumb-to-Finger Microgestures. In *Proceedings of the 2018 ACM International Conference on Interactive Surfaces and Spaces*. ACM, 177–187. DOI: `http://dx.doi.org/10.1145/3279778.3279799`

[25] Hyunyoung Song, Hrvoje Benko, Francois Guimbretiere, Shahram Izadi, Xiang Cao, and Ken Hinckley. 2011. Grips and gestures on a multi-touch pen. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 1323–1332. DOI: `http://dx.doi.org/10.1145/1978942.1979138`

[26] Brandon T Taylor and V Michael Bove Jr. 2009. Graspables: grasp-recognition as a user interface. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 917–926. DOI: `http://dx.doi.org/10.1145/1518701.1518842`

[27] Margarita Vergara, JL Sancho-Bru, V Gracia-Ibáñez, and A Pérez-González. 2014. An introductory study of common grasps used by adults during performance of activities of daily living. *Journal of Hand Therapy* 27, 3 (2014), 225–234. DOI: `http://dx.doi.org/10.1016/j.jht.2014.04.002`

[28] James A Walsh, Stewart Von Itzstein, and Bruce H Thomas. 2014. Ephemeral interaction using everyday objects. In *Proceedings of the Fifteenth Australasian User Interface Conference-Volume 150*. Australian Computer Society, Inc., 29–37. `http://dl.acm.org/citation.cfm?id=2667657.2667661`

[29] Raphael Wimmer. 2011. Grasp sensing for human-computer interaction. In *Proceedings of the fifth international conference on Tangible, embedded, and embodied interaction*. ACM, 221–228. DOI: `http://dx.doi.org/10.1145/1935701.1935745`

[30] Raphael Wimmer and Sebastian Boring. 2009. HandSense: discriminating different ways of grasping and holding a tangible user interface. In *Proceedings of the 3rd International Conference on Tangible and Embedded Interaction*. ACM, 359–362. DOI: `http://dx.doi.org/10.1145/1517664.1517736`

[31] Robert Xiao, Julia Schwarz, Nick Throm, Andrew D Wilson, and Hrvoje Benko. 2018. MRTouch: adding touch input to head-mounted mixed reality. *IEEE transactions on visualization and computer graphics* 24, 4 (2018), 1653–1660. DOI: `http://dx.doi.org/10.1109/TVCG.2018.2794222`

[32] Dongwook Yoon, Ken Hinckley, Hrvoje Benko, François Guimbretière, Pourang Irani, Michel Pahud, and Marcel Gavriliu. 2015. Sensing tablet grasp+ micro-mobility for active reading. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology*. ACM, 477–487. DOI: `http://dx.doi.org/10.1145/2807442.2807510`

[33] Joshua Z Zheng, Sara De La Rosa, and Aaron M Dollar. 2011. An investigation of grasp type and frequency in daily household and machine shop tasks. In *2011 IEEE International Conference on Robotics and Automation*. IEEE, 4169–4175. DOI: `http://dx.doi.org/10.1109/ICRA.2011.5980366`