VirtualGrasp: Leveraging Experience of Interacting with Physical Objects to Facilitate Digital Object Retrieval

Yukang Yan[1], **Chun Yu**[123†], **Xiaojuan Ma**[4], **Xin Yi**[1], **Sun Ke**[1], **Yuanchun Shi**[123]
[1]Department of Computer Science and Technology
[2]Key Laboratory of Pervasive Computing, Ministry of Education
[3]Global Innovation eXchange Institute, Tsinghua University, Beijing, 100084, China
[4]Hong Kong University of Science and Technology, Hong Kong
{yyk15,yix15,k-sun14}@mails.tsinghua.edu.cn {chunyu,shiyc}@tsinghua.edu.cn mxj@cse.ust.hk

Figure 1: Examples of the object-gesture mappings designed by users under the metaphor of "grasping the objects". Users adapt their hand postures to different physical objects that they grasp. The objects from left to right: Toy gun, mug, book, stapler, phone, and pen.

ABSTRACT
We propose VirtualGrasp, a novel gestural approach to retrieve virtual objects in virtual reality. Using VirtualGrasp, a user retrieves an object by performing a barehanded gesture as if grasping its physical counterpart. The object-gesture mapping under this metaphor is of high intuitiveness, which enables users to easily discover, remember the gestures to retrieve the objects. We conducted three user studies to demonstrate the feasibility and effectiveness of the approach. Progressively, we investigated the consensus of the object-gesture mapping across users, the expressivity of grasping gestures, and the learnability and performance of the approach. Results showed that users achieved high agreement on the mapping, with an average agreement score [35] of 0.68 (SD=0.27). Without exposure to the gestures, users successfully retrieved 76% objects with VirtualGrasp. A week after learning the mapping, they could recall the gestures for 93% objects.

ACM Classification Keywords
H.5.2. [Information Interfaces and Presentation]: User Interfaces: Input devices and strategies

†donates the corresponding author.

CHI 2018, April 21–26, 2018, Montreal, QC, Canada
© 2018 ACM. ISBN 978-1-4503-5620-6/18/04…$15.00
DOI: https://doi.org/10.1145/3173574.3173652

Author Keywords
Gesture; Object selection; Mapping.

INTRODUCTION
Virtual Reality is growing to be an important platform for various types of applications, including games [18], training [20] and educations [27]. Object Retrieval is a common and basic task in these applications. For example, a game player often picks up weapons and a mechanic might need several tools (e.g. a spanner). Currently, two main ways for users to access virtual objects are to pick them from the scene or to pick them out from a menu. The typical interfaces of these two approaches in popular VR games are shown in Figure 2. In both cases, users need to specify the position of the target with a device (e.g., a controller) or their finger [38]. However, to pick up the target from the scene may encounter several problems, especially when the objects are occluded, placed in a very dense layout or when the controller jitters reduce the accuracy. To select an item in a menu requires a series of manipulations. Users need to invoke the menu, choose the category, scan the items in the sub-menus, until they pinpoint the desired one and get back to the ongoing task. The manipulations could be time-consuming and distracting, especially when the users are new to the interface or when the target is buried deeply in a hierarchical menu [12].

Compared to these pointing and menu techniques, retrieving virtual objects with the assigned gestures could avoid the positioning problem and simplify the searching process. A key to the design of the gesture-based object retrieval is the mappings between objects and gestures, which greatly influence the learnability of the approach. A good mapping design

should satisfy several criteria. It should be easy to discover and memorize [34, 47], be consistent with the acquired experience of users [30], and gains high consensus across users [48]. Most previous work on gesture input was developed to issue commands, which leaves the mapping for retrieving objects to be studied.

Figure 2: The menu interface in "Destinations" [2], where users first select "Furniture" in the "Things" menu (top left) and select a mug in the sub-menu (top right); the scene in "RecRoom" [6], where users access the position of the prop to pick up it (bottom).

In reality, prehension (i.e., the action of grasping with hand and fingers) is one of the most fundamental elements of humans' physical interaction with an object, and the grip of an object must be adapted to its shape, size as well as intended use [35], as illustrated in Figure 1. Inspired by this fact, we propose a novel gesture-based approach VirtualGrasp, to retrieve virtual objects with the grasping gestures. To retrieve an object with VirtualGrasp, users perform a static barehanded gesture in the air as if grasping its physical counterpart. For example, users can perform a gun-holding "hook" gesture to retrieve a virtual gun. In this manner, the objects themselves remind users the gestures for retrieval which is also consistent with users' own experience. As it is a fact that not all general objects are graspable, e.g. too large objects or abstract objects, this approach aims to enable the retrieval of everyday graspable objects in VR applications.

However, are the object-gesture mappings consistent across different users? Do the grasping gestures of different objects confuse with each other? Is it easy to discover and learn the mappings? To answer these questions, we conducted three studies and their workflow is visualized by Figure 3. Study1 is a gesture elicitation study to probe the consensus of the mappings across users; Study2 is a gesture classification study to measure the expressivity and confusions of the grasping gestures; Study3 is an object retrieval study to evaluate the performance and learning effort of users with VirtualGrasp. The results showed that users achieved good agreement on the mappings and the algorithm we developed could correctly recognize a majority of the desired objects (37/49) of users. Users easily discovered, learned and recalled the object-gesture mappings and the approach was well-accepted. We demonstrate that VirtualGrasp could promote user experience when applicable and has potential to supplement pointing and menu techniques for virtual object retrieval.

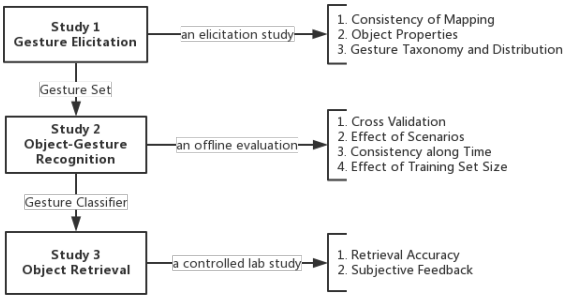

Figure 3: The workflow of the three studies, showing the input, output, and relationship of the studies.

RELATED WORK

Gestural Object Acquisition

In 3D space, acquiring object's position can be achieved by two main approaches, which are virtual pointing and virtual hand [9]. Virtual pointing techniques enable users to emit a ray to point at the object, from their hands or the controller [16, 22, 42, 51]. Virtual hand techniques enable users to directly touch the object with the controller or bare hand to acquire it [38, 44, 53]. Besides, users can also retrieve objects from menus [13, 17, 21] by searching the item from the lists.

However, in this paper, we proposed a gesture-based approach to retrieve objects in VR, VirtualGrasp. Gesture-based input is a category of interaction modality in which users perform static or dynamic gestures with their hands, arms or other body parts to input information to computers [26]. Comparing to the approaches mentioned above, the gesture-based approach has the advantage of enabling bare-hand and eyes-free and direct input [10]. Performing a gesture does not require users to visually search the desired object or switch to another interface to search the menus. However, the available gestures for a gesture-based interface are not explicitly displayed and users have to learn the mapping of gestures and commands in a manual. How to help users discover and remember the gestures deserves further study.

Gesture-Object Mapping

A key to the design of gesture-based systems is the mapping between targets and assigned gestures. The mapping, to some extent, determines the discoverability and learnability of the gestures [25, 47]. However, in some of the current systems [31, 33, 49, 50], the primary goal of the system designers is to design the gestures to be robust to recognize, instead of to design the mappings to be intuitive to the users.

In previous research, there are mainly two ways to address the learnability issue. One way is to specially design the look-and-feel of the target to suggest the assigned gesture. The gestures were mapped to the shapes, colors, motions of the targets, or directly overlaid onto the targets [14, 15, 19, 30, 52]. Escape [52] mapped directional swipe gestures to the icons with different colors, shapes, and patterns. Orbits [19] and Pathsync [15] enabled users to select targets on smart watches and remote displays by following the movements of the targets with their gaze and hands respectively. Gesture Select [14] directly overlaid the stroke onto the targets, which users draw with a mouse. Finger-Count menus [47] overlaid different numbers onto the targets, which were according to the number of the fingers users need to stretch out to select them. In these cases, the gestures were cued by the appearance of the targets, and thus are easier to discover and remember [52]. However, using these techniques, users need to observe the targets to find the gestures, which requires a high visual load.

The second but more widely used solution is the user-defined approach, which was first introduced by Wobbrock et al. [48] when designing gestures on an interactive surface. This approach involves end users into the design process of the gesture-command mappings. They portray the effect of a command, e.g. to delete an item, and then ask a group of users to design their own gestures to issue this command. The gestures with the highest consensus will be assigned to the commands. As a result, the elicited command-gesture mappings reflect daily behaviors and experience of users, which results in a more contextual connection between gestures and commands [48]. The approach has been successfully applied to many areas [8, 29, 40, 45]. However, for each command, users might prefer different gestures to assign to it, which were not all supported by the gesture set [8, 29].

For these two ways, our research is more relevant to the user-defined gestures. However, our focus is the retrieval of graspable *objects*, which is very different from the *commands* that previous research studied. As users have experience of interacting with the physical objects, we expect the gesture-object mappings designed by users to be more consistent with each other, and therefore easier to be discovered and remembered by users.

Imaginary Interaction

Another body of related research is the *Imaginary Interactions*, which share the idea to transfer user experience into HCI. We summarize a series of imaginary interactions, where users replicate the interaction process with physical objects onto imaginary objects. Imaginary Interface [23] enabled users to interact with an interface that existed in their own imagination. After forming an L-shaped coordinate cross with the non-dominant hands, users could point and draw accurately in this origin of imaginary space. Imaginary Phone [24] enabled users to perform micro-operations of a smart phone by tapping and sliding at the relative locations on their empty hands. Imaginary Devices [43] enabled users to do the pointing tasks in games by mimicking the operation of the joystick, the steering wheel or other control devices. Imaginary Reality Gaming [11] enabled users to play with others in a ball game using an imaginary ball. Users learned the position of the ball by watching each other act and the occasional auditory feedback.

In this paper, our contributions compared to their research include: 1) We transfer the concept from using specific objects, e.g. a phone or a ball, to the retrieval of a wide range of objects that users could grasp with their hands; 2) We fill in the gap to test the consensus of the object-gesture mappings before applying them for retrieval tasks.

Grasping Objects

Before this paper, several studies in HCI domain [46, 36, 25] have discussed the strong connection between objects and the gestures to grasp them. One study [36] tried to detect the physical objects that the user is interacting with in an office by recognizing the hand gestures. They found that users performed different gestures even when interacting with the same object. This made it difficult to build one general model for all users. Another study [46] explored to detect the size and shape of the objects by the grasping gestures of users. They tested objects with three levels of sizes and six basic shapes. The result showed that the grasping gestures could help discriminate objects with different sizes and shapes.

There is work that demonstrated to leverage grasp gestures to select tools on touchscreens, TouchTool [25]. They enabled users to select the tools by mimicking the grasping gestures of them onto a 2D touchscreen. They evaluated the intuitiveness and convenience for input on seven specific computer tools, e.g. mouse and camera. The number of tools was limited by gesture conflicts. Compared to TouchTool, we move the surface-based 2D gesture space to the 3D space in the air. Without the constraints of the touchscreen, we expect fewer conflicts between gestures and higher intuitiveness of the gestures. As a result, we aim to enable users to retrieve a larger number of objects.

VIRTUALGRASP

In reality, users have sufficient experience of interacting with physical objects. A fact is that the gestures users use to grasp or manipulate objects are adapted to the different shapes, sizes and intended uses of the objects. For example, the "hook" gesture we use to grasp a mug is adapted for its ring-shaped handler; the "ten-finger-typing" gesture is adapted for the keyboard layout and purpose of fast typing. We termed these gestures to be the "*grasping gesture*" in this paper when no ambiguity can arise.

Based on this fact, we propose VirtualGrasp, an approach that enables retrieving virtual objects by performing grasping gestures of their physical counterparts. There are several benefits that VirtualGrasp could possibly provide. First, as the objects themselves remind users of the grasping gestures, we expect that VirtualGrasp requires little learning effort from users. Once users know the "grasping" metaphor, we expect the gestures of VirtualGrasp to be self-revealing to them. Second, VirtualGrasp does not require users to switch to the menu interfaces or visually search the positions of the targets, thus this enables users to continuously focus on the ongoing task and causes fewer distractions.

Hypotheses

However, there are three hypotheses to be tested before we could achieve these benefits:

1. Consensus: Users achieve high agreement on the mappings of objects and the grasping gestures.

2. Expressivity: A good number of grasping gestures can be correctly recognized and distinguished by algorithms.

3. Self-revealing: Users can discover the grasping gestures of the desired objects by themselves or it is easy to learn and remember the gestures.

To test these hypotheses, we conducted the following three studies. As shown in Figure 3, through the studies, we completed the end-to-end exploration of the VirtualGrasp design. We went through a gesture elicitation study, the development and evaluation of the gesture classifier, and an object retrieval study to evaluate the user performance using VirtualGrasp. After the studies, we discuss the design implications and limitations of our approach.

STUDY1: GESTURE ELICITATION

The goal of this study is to probe the consensus of the gesture-object mappings defined by users. We first generated an object list through a brainstorming session. Then we invited participants to design the grasping gestures for the objects in the list. Based on the obtained object-gesture pairs, we measured the consistency of the mappings, extracted the important properties of objects that helped design the gestures, and analyzed the distribution of gestures in a taxonomy.

Generation of the Object List

In previous gesture elicitation studies [8, 29, 40, 45, 48], researchers generated a list of most commonly used commands, and then invited users to design gestures to issue them. However, very different from commands, everyday graspable objects cannot be all included in this elicitation study. Therefore we need to generate a object list that could represent the countless graspable objects. To be representative, the list should cover the diversity of different shapes, sizes and include enough number of objects. Meanwhile, the number of the objects should not be too large for this study.

To obtain this list, we conducted a brainstorming session with 26 participants (10 females, aged from 23 to 26). All the participants were graduate students from a local campus. Eight of them had VR experience. All of them were familiar with touch-screen gesture interaction. To avoid that users come up with the objects that could not be grasped, e.g. too large objects, we first explained the basic idea of VirtualGrasp, which was retrieving the objects with their grasping gestures. Meanwhile, we deliberately instructed them to not consider the performance of gesture recognition. Then we asked participants to list the objects they could retrieve with VirtualGrasp.

In total, we collected 215 votes which contained 101 distinct objects. Among them, 43 objects received only one vote, "camera" received the most (15) votes. To limit the number of objects for user study, we removed objects that received no more than three votes (52 objects (62 votes): 43 $\times$ 1 vote, 8 $\times$ 2 votes, 1 $\times$ 3 votes) and meanwhile kept the object list representative. We finally generated a list of 49 objects, which represented 153 votes, which maintained more than 70% of the 215 votes. We also confirmed that they could cover the diversity of different shapes and sizes. We went through the merged object list with participants and split them into six groups that could appear in the same scenarios.

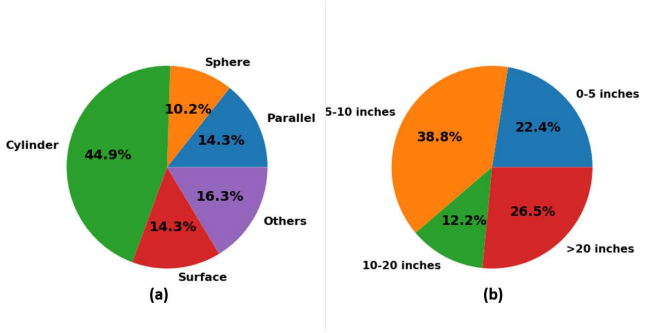

Figure 4: The distribution of the objects on the (a) shape and (b) size.

We analyzed the distribution of the object list on the shape and size. We applied an existing taxonomy [46] to categorize the shape of the object. We measured the size by the largest dimension among the length, width, and height of the object [1, 7]. As Figure 4 shows, the objects in our list covered basic shapes and sizes of everyday graspable objects.

Participants

We recruited 20 participants (14M/6F) from a local campus, with an average age of 23.6 (from 20 to 27). Eight of them participated in the brainstorming session. In a pre-study questionnaire, participants reported their familiarity with six scenarios in a 5-point Likert scale. The average scores were 4.9, 2.8, 3.7, 4.8, 4.6 and 4.8 in the order same as Table 1. Ten participants had experience of mid-area gesture interaction, using LeapMotion or Kinect. All the participants were right-handed.

Design

The factor of this study was the object we presented to the participants, for which they would design a grasping gesture. For each object, we recorded the grasping gestures participants designed by taking pictures. To log the three-dimensional information of hand gestures, we took pictures from a front and side view of participants performing the gestures. Consistently with previous studies [29, 40, 45, 48], we instructed users to focus on the gesture recall and assume all the conceived gestures could be recognized correctly. Each time after participants designed a gesture for each object, we asked them which type of object properties helped them recall the gestures. The properties included shape, size, texture, temperature, location, orientation, motion and usage. To inform participants the object, we showed the names of the objects instead of pictures or other visualizations. The consideration was that we expected VirtualGrasp to also support retrieval of out-of-view objects. All the participants stood during the experiment.The experiment was conducted in a quiet office room.

Scenarios	Object lists
Office	book, briefcase, eraser, mouse, keyboard, pen, scissor, stapler
Game Weapons	binocular, bow, dagger, grenade, handgun, rifle, shield, spear, sword
Sports	barbell, basketball, badminton racket, cue, golf club, javelin, ski stick, shot, skipping rope
Electronics	camera, flash drive, headphone, interphone microphone, phone, remote control
Home	bowl, broom, comb, glasses, mug, perfume, toothbrush, umbrella, watch
Food	apple, banana, beer, hamburger, popsicle, watermelon

Table 1: The object list for the study, which was divided into six groups, and objects in the same group could appear in the same scenarios.

Procedure

First, we introduced the task, which was to design a gesture when the target object was informed. We instructed participants to recall the gestures that they used to grasp the target object in reality. Each participant performed 49 trials of gesture design. In each trial, participants first saw the name of the target object on the front screen. Then they were given time to recall the gesture. After they decided, they performed the gesture with either or both of their hands to define it. At the same time, the experimenter took pictures of the gesture. Then they were asked to tick on the object properties on a questionnaire that helped them recall the gesture. The participants were free to rest when they felt tired. The experiment took about half an hour on average for each participant.

Data Processing

From 20 participants, we collected 980 (20×49) object-gesture pairs. Before further analysis, we observed that some of the pairs seemed to be unreasonable. An example was that only one participant gestured a shot by lifting both hands over head, which was composed as lifting one hand beside the check for the other 19 participants. As the composer reported, the unique gesture was created based on his imagination because he never threw a shot or watch shot puts before. To filter out these unreasonable pairs, we conducted a voting session with 40 participants (20 extra participants), where they voted for top three gestures for each object that they thought to be most matched. We found 39/980 pairs without any votes. These pairs were most in Sports (15) and Game Weapon(12) scenarios where participants lack interaction experience.

To be consistent with previous elicitation studies, we kept these pairs in the consistency analysis. However, as the composers reported, these pairs were mostly created by their own imagination and the mappings were not intuitive to others. So we decided to remove them from our final gesture set and the following studies. We merged the gestures for each object. We shared a similar procedure with [8], in which two of the authors coded gestures displayed in photos. The criteria for gesture classification included single/double hands, the position, orientation and shape of the hands. Details of the criteria are shown in Table 2. We finally generated 140 distinct object-gesture mappings.

Result

Based on the 980 object-gesture pairs in 140 categories, we measured the consistency of participants' mappings, extracted the key object properties and analyzed the distribution of the gestures based on a taxonomy.

Consistency of Mappings

We applied two metrics to evaluate the consistency of the object-gesture mappings across participants. The first metric was the number of different gestures mapped to each object. Under the criteria of gesture classification, we found that a large portion (18/49) of the objects were mapped to one unique gesture that all the participants agreed on. Besides, all of the objects were mapped to no more than five gestures.

The second metric we applied was the agreement score proposed by Wobbrock [48], for characterizing the level of consensus between participants' proposals. The agreement score for a given referent r (a target object) for which feedback gestures were elicited from multiple participants was calculated by the following formula:

$$A(r) = \sum_{P_i \subset P} \left(\frac{|P_i|}{|P|}\right)^2 \quad (1)$$

Where P is the set of all proposals for referent r, P_i is the subset of identical proposals from P. Here is an example to better understand the metric: Two groups of 10 participants proposed two different gestures for object A, and the agreement score was calculated to be 0.50; meanwhile, two groups of 1 and 19 proposing two gestures for object B resulted in 0.905. Both object A and B were mapped to two different gestures, but the consistency of the mappings was very different. In our study, the average agreement score of 49 objects was calculated to be 0.68 (SD=0.27), which was much higher than previous gesture elicitation studies [48, 8, 29, 40, 45]. 36 of 49 objects achieved a score of equal or more than 0.50, which could be regarded as indicators of robust proposals [45]. If we removed the 39 object-gesture pairs that were voted to be unmatched in Data Processing, the measured consistency would be higher (score = 0.70). Both metrics indicated that participants could reach high agreement on the mapping of objects and the grasping gestures.

Key Object Properties

Participants reported key properties that helped them recall the grasping gesture for each object. We listed eight different properties: shape, size, texture, temperature, position, orientation, motion and usage. The result showed that on average, 41.3/49 and 40.8/49 objects were mapped to gestures according to their shapes and usages, and 29.8/49 objects were suggested by their sizes. Other properties had an influence on less than ten objects on average. The results revealed that the intended usage of the object played an important role in the mappings of objects and gestures. Even two objects with similar shapes and sizes, e.g. a snooker cue and a javelin, were mapped to

very different gestures (double-hand horizontally holding gesture vs single-hand throwing gesture) because of the different usages. However, for some objects, the shape and size were exactly designed to serve its potential use, but we did not analyze this interaction effect in this study.

Gesture Taxonomy and Distribution

We present a taxonomy of the elicited gestures, mostly based on their physical formation. Previous studies also classified gestures by their "Nature" dimension, which defines the mapping of gestures to the commands [48]. While all of our gestures fell in the "Physical" level, as the gestures were used to perform on physical objects. Our taxonomy consisted of four dimensions, which were "Single/Double Hands", "Hand Position", "Palm Orientation" and "Hand Shape". "Hand Position" referred to the hand position relative to the main body, e.g. in front of the body; "Hand Shape" was categorized into seven classes, based on an existing taxonomy for hand postures described by Mackenzie and Iberall [46]. However, we found some gestures, e.g. a planed-hand pressing gesture for "Stapler", could not be mapped to their six basic categories, so we placed them at the level of "Others". More details of each dimension of the taxonomy are listed in Table 2.

Figure 5 visualizes the breakdown of our gestures. As shown, single-handed gestures were the majority (61.4%). Both in single and double hands gestures, the most frequent hand shape was "Cylindrical" (45.4%, 44.3%), followed by "Palmar" (24.4%, 16.6%). The most frequent hand orientation was the orientation facing the body. The number of different levels in our taxonomy was 141, which was close to the number of elicited object-gesture pairs. From this distribution, we could also dig out some gestures formed by the features in infrequent levels in the dimensions, e.g. a *Tip* shaped hand in the *Side* area with an *Upward* orientation. These gestures had a small possibility to conflict with the gestures of grasping objects. So designers could apply them to issue commands when they concern about avoiding the false positives.

Discussion

The experimental results supported our hypothesis that by simply providing the metaphor of "grasping the objects", participants could reach high agreement on the mappings between objects and the grasping gestures. Compared to previous elicitation studies, we applied a half open-ended design as we told the users this metaphor. The results showed that this design led to a much higher agreement score of the mappings, and also maintained the intuitiveness of the mappings.

In previous elicitation studies, the final gesture set only contained one gesture with highest votes for each command. However, as our taxonomy suggested, there was potential to classify 141 different gestures correctly. After we removed 39 gesture-object pairs that were voted to be unmatched, our final gesture set contained 101 pairs. We saw an opportunity to support multiple gestures for each object, which may meet users' first intuition with higher possibility and alleviate their learning cost. So in Study 2, we decided to develop an algorithm to classify all 101 object-gesture pairs.

Hand	Single	The gesture was performed with one hand
	Double	The gesture was performed by both hands
Orientation	Forward/ Back/ Up/ Down/ Inner	Single-hand gesture that the hand faces to the each direction
	Faced	Both hands face to each other
	Same	Both hands face to the same direction
Position	High	The hand(s) was/were higher over shoulders
	Side	The hand(s) was/were at the side of the body
	Front	The hand(s) was/were at the front of the body
Posture	Cylindrical	Open fist grip used for grasping tools or a closed fist for thin objects
	Spherical	Spread fingers and arched palm to grasp spherical objects
	Tip	Fingers grasp sharp and small objects, such as a needle or pen
	Hook	Used for heavy loads
	Palmar	Used for flat and thick objects
	Lateral	The thumb is primarily used in order to grasp thin and flat objects such as a piece of paper

Table 2: The detailed dimensions of the taxonomy, which was also the criteria for classifying the gestures.

For the 39 object-gesture pairs that we removed in the final gesture set, we confirmed with the composers that they had little experience of using these objects. The lack of experience resulted in arbitrary mappings based on their imagination. This phenomenon revealed that the mappings would also be affected by the background and experience of the users. In the future, we can invite more users with a higher variety of professions and backgrounds to test this effect.

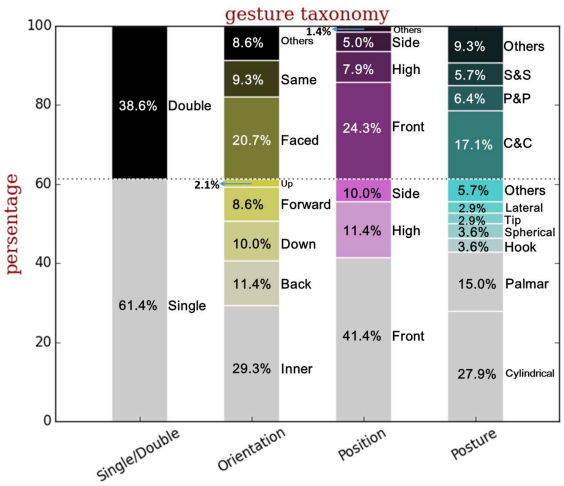

Figure 5: Distribution of the gestures that users designed in this study in four dimensions of the taxonomy.

STUDY2: ACCURACY OF GESTURE RECOGNITION

In this study, we aimed to develop an algorithm that can recognize the grasping gestures and predict the corresponding objects, which mappings were obtained in Study 1. We first collected gesture data from users, and then trained a linear

SVM model and implemented the classification algorithm based on it. At last, we evaluated the classification performance through offline validations.

Participants

We recruited twelve participants, with an average age of 24.3 (SD = 1.5) in this study. All the participants were recruited from a local campus. Four of them had experience of mid-air gesture interaction. All of the participants were familiar with touchscreen gestures and were right-handed. Seven of the participants participated in Study 1.

Figure 6: The sensors that participants put on to record the hand gesture data, on finger joints, hand backs, arms and the head. They put on the sensors before the experiment with the help an experimenter.

Apparatus

We used a MEMS (Micro-Electro-Mechanical System) tracking device, Perception Neuron to record the hand gesture data of participants. It was an inertial sensor-based tracking device, with a resolution of 0.02 degree. Participants need to put on the sensors at the appointed positions before the experiment. We developed a program to show participants the gestures they should perform, using Unity 5 engine. The program showed the pictures of users performing the gestures in Study1.

Procedure

Participants first put on the sensors of Perception Neuron, as shown in Figure 6. Then they started two rounds of 101 trials of tasks, in which they were required to perform the gesture that were informed by an picture from the front view. In each trial, participants first observed the picture to understand how to perform the gesture, shown on the front screen. Then they were given time to practice. They were told to pay attention to single/double hands, hand position, palm orientation and hand shape of the gesture. Then they performed the gesture and dwell for 1 second, and meanwhile, the program recorded the gesture data. The order of the tasks in each round was randomized and the participants had a break between two rounds. The whole process took around one hour on average.

Data Processing

We recorded positions of 14 joints of five fingers relative to the position of the palm, the position and orientation of each hand palm for each gesture, for 40 frames. Data of each frame for each gesture formed a vector of 96 values (2 hands $\times$ 16 vectors $\times$ 3values). For the single-handed gestures, participants put the other hand down in the rest region, and we set the corresponding values to be zeros. We obtained 12 participants $\times$ 101 gestures $\times$ 2 rounds $\times$ 40 frames = 96960 frames of data in total. We then implemented a supervised machine learning classifier (SVM). We trained the classifier with the collected data, where the feature vectors of the gestures were the input and the corresponding objects were the output.

To evaluate the classification performance, we calculated Top-1, Top-3, Top-5 accuracy as the metrics. A metric of Top-x accuracy was calculated as the frequencies that x most possible objects, predicted by the classifier, contained the correct objects for the input gestures. Top-1 accuracy was an intuitive metric which represented the possibility that our algorithm could return the exact object that users perform gestures to retrieve. We also applied Top-3 and Top-5 accuracy because they would also be valuable if allowing users to perform an extra selection among these three or five object candidates.

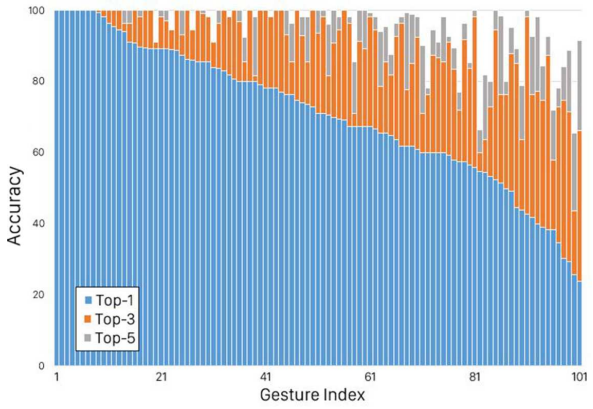

Figure 7: Average classification accuracy for 101 object-gesture pairs in Leave-Two-Out cross-validation, measured by Top-1, Top-3 and Top-5 accuracy in different colors.

Results

We evaluated the classification accuracy through a series of offline tests. We cross-validated the performance for held-out participants, within different scenario groups and with the training sets of different sizes.

Leave-Two-Out Validation

Leave two subjects out cross-validation was performed, where we held out the data of every two participants to be the test set and use the rest data to be the training set. 66 tests (C_{12}^2) were performed and we calculated the Top-1, Top-3, Top-5 accuracy. Figure 7 visualizes the average accuracy of 101 object-gesture pairs, in the descending order of Top-1 accuracy. The average accuracy of all 101 pairs was calculated to be 70.96% (SD=9.25%) for Top-1, 89.65% (SD=6.39%) for Top-3 and 95.05% (SD=4.56%) for Top-5. Eight pairs had 100% Top-1 accuracy and the lowest Top-1 accuracy was 31.5%, which was the grasping gesture for "pen". As the performance had a big difference among object-gesture pairs, we extracted 20 most frequently mismatched pairs to probe the reasons for the confusions. We found that the grasping

gestures for very small objects (flash drive, pen, eraser) were easily confused, which may be because that users grasp them roughly in a very similar way. In these cases, the size of the objects dominated the grasping gesture and the gestures hardly reflected the object shapes or intended usages. We also found that the usage of the objects played a strong part in the object-gesture pairs with high classification accuracy (100%). These gestures seemed unique to the objects, such as stapler - the pressing gesture, basketball- the shooting gesture, glasses - the gesture of putting on them. Therefore these gestures were also easier to be distinguished from others.

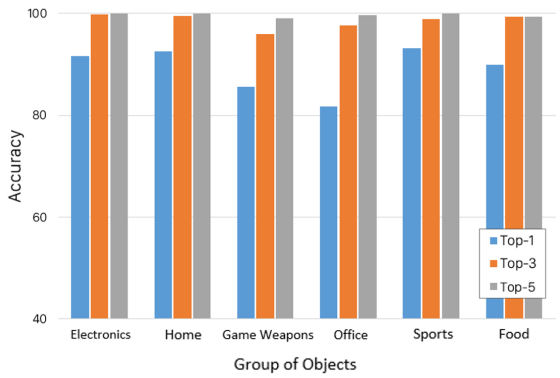

Figure 8: Average classification accuracy within six scenario groups, measured by Top-1, Top-3, and Top-5 accuracy.

Effect of Scenarios and Training Set

We tested the performance of classifying the object-gesture pairs in the scenarios that they might appear. By leveraging the context information of the scenario, some of the gesture confusions could possibly be avoided. For example, a pen is more possible to appear in an office than a toothbrush, although their grasping gestures look similar. We grouped gesture data by the scenarios that the corresponding objects belonged to. We performed *Leave-Two-Out* validation again within each group, and the result was shown in Figure 8. Top-1 accuracy for all six scenarios was higher than 80% and Top-3, Top-5 accuracy were all over 95%. Compared to the classification of all 101 object-gesture pairs, the performance was improved after considering the effect of scenarios. Comparing the performance of six scenarios, we found the object-gesture pairs in "Sports" had the best performance.

We also tested the performance change of classification when the size of the training set was different. We initiated the training set with the data of P1 and enlarged it by the data of another participant each time, and we fixed the testing set to be the data of P12. We performed a Pearson's test to measure this correlation. The result showed that there was a very strong linear correlation between the accuracy and training set size ($r=0.975, 0.920, 0.946$; $p<0.0001$). We also split the gesture data of participants by two rounds and then we tested the consistency between two rounds. We set the training set to be the data of first round and the test set to be the second round. The accuracy were calculated to be 82.92% (SD=6.85%) for Top-1, 98.27% (SD=1.53%) and 100% (SD=0). Compared with Leave-Two-Out validation, the result showed that a larger training set or recording the gesture data of new users could improve the classification performance.

STUDY3: RETRIEVE OBJECTS WITH VIRTUAL GRASP

In this study, we evaluated the performance of retrieving objects using VirtualGrasp in a VR application. There were two hypotheses to test in this study: 1. After knowing the "grasping" metaphor, Users could discover the object-gesture by themselves; 2. Users could easily remember and recall the object-gesture mappings. Besides, we also collected the think-aloud comments participants made and their subjective feedback.

Participants

We recruited twelve new participants who did not participate in Study 1 or Study 2. Therefore, they did not know any of the object-gesture mappings before this experiment. These participants (8M/4F) were aged from 21 to 25 (AVG=23.1). All the participants were familiar with touchscreen gesture interaction on smart phones. Three of them had experience of mid-air gesture interaction. All participants used their right hand as the dominant hand.

Apparatus

We implemented the experiment platform, which is shown in Figure 9. The name of the target object was shown on the top; the object candidate list predicted by the algorithm was visualized in the center after the gesture is recognized; the hands and arms of participants were visualized in the bottom to help them adjust their gestures. Participants wore the same tracking sensors, Perception Neuron to input gestures. The experiment was conducted in a quiet office.

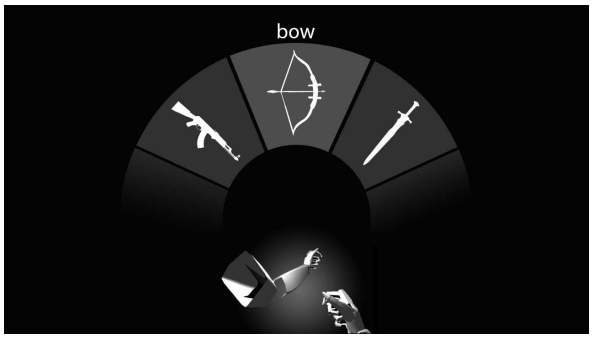

Figure 9: The user interface of the experiment, showing the task (bow), object candidates (bow, rifle, and sword) and current gesture (the visualization of arms and hand).

Algorithm

We used the same data format of Study2 to record participants' input gestures, which was a feature vector of 96 values (2 hands $\times$ 16 positions $\times$ 3 values). We implemented the gesture classifier based on a linear SVM, trained with all gesture data of Study2. The classifier received the feature vector of the gesture as the input gesture and calculated the possibility that it was mapped to each object, and finally returned a list of 49 possibilities. For different grasping gestures of one

object, we calculated the highest possibility among them to be the possibility of the object. We showed participants the three objects with the highest possibilities as the candidates to choose from.

Design and Procedure

This experiment consisted of three sessions of 49 trials of object retrieval. In each trial, the task of participants was to perform a gesture according to the target object. They were instructed to perform the gesture that they used to grasp the object in reality. A dwell time of 0.5 seconds was regarded as a commitment of the gesture. When not performing gestures, participants were instructed to put both hands down in the rest region. After the gestures were recognized, three most possible objects were returned. Participants performed directional swipes to choose among them. When the target object was not included in the list, participants directly skip to the next trial.

We arranged three sessions along time: the discovery, learning, and recall session. Participants completed the discovery and learning session during the first time they came, whereas they came back and completed the recall session a week later. In the discovery session, participants performed their own grasping gestures according to the target objects. Without learning of the object-gesture mappings in the systems, they were encouraged to perform the gestures of their first intuition under the metaphor of "grasping objects" and not to guess the standard gestures in the system. In the learning session, participants learned the standard gestures in the systems, before using them to retrieve the objects. We showed the pictures of the hand gestures and participants were free to practice them before they finally chose one to perform. In the recall session, participants were instructed to recall one of the gestures they learned in the learning session and perform it to retrieve the target object. If they forgot all the gestures for certain object, they still performed their own gestures. In the duration of one week, we ensured that they were not exposed to the standard gestures. The order of target objects was randomized in each session. On average, the discovery and recall session took fifteen minutes and the learning session took half an hour to complete. After three sessions, participants filled in a questionnaire about their subjective feedback and comments.

Results

Classification Accuracy

We measured user performance with the metrics of Top-1, Top-3, Top-5 accuracy, and Figure 10 shows the average performance for three sessions. In the discovery session, participants could discover the exact gestures for target objects with a proportion of around 40%. Within a list of five possible objects, the accuracy increased to 75.51%. This result proved the self-revealing feature of VirtualGrasp. With VirtualGrasp, users without any training could retrieve 37/49 virtual objects successfully, which was hardly achieved by previous systems. After learning, the performance increased to 59.35% (Top-1), 89.46% (Top-3) and 94.5% (Top-5). Compared to the results in Study2, the Top-3 and Top-5 accuracy were comparable but the Top-1 accuracy was lower. This was possibly because all 101 object-gesture pairs were not balanced in the choices of the users. For each object, they might have chosen the gesture that they felt most intuitive but there might be another gesture that was easier to recognize and led to higher accuracy. In the recall session, the performance had a modest reduction from the learning session, which was 55.78% (Top-1), 86.39% (Top-3) and 93.20% (Top-5). However, it was still a great promotion from the discovery session. The results in all three sessions showed that the mappings of VirualGrasp were easy to discover, memorize and recall.

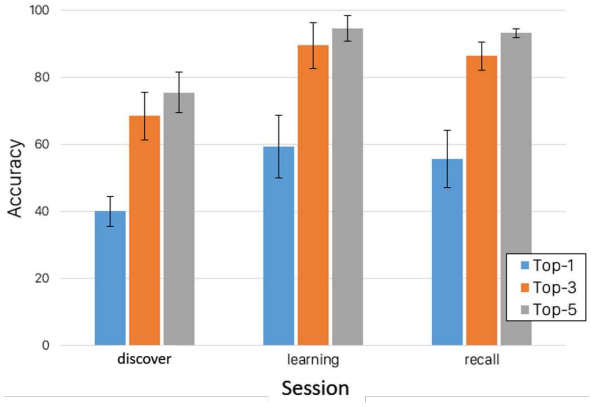

Figure 10: The average classification accuracy of 49 objects in the discovery, learning and recall sessions, measured by Top-1, Top-3 and Top-5 accuracy.

Subjective feedback

We report the comments that participants made during the experiment, which could reflect their experience while retrieving objects with VirtualGrasp.

In the discovery session, some participants first got confused when they were asked to perform the grasping gestures, but surprised when the desired object was shown in the candidate list. *"Two different gestures came to me for grasping the camera and it was intelligent that the system correctly recognized the one I performed."* [P4]

In the learning session, for most participants, they could agree with the supported object-gesture mappings, especially for the object that they both hardly use in reality and also failed to retrieve in the discovery session. *"I never used a grenade before, but I agreed with Gesture 3 of grasping it over the shoulder to throw it."* [P6]

In the learning and recall session, after learning the whole set of object-gesture mappings, some participants found tricks to improve the accuracy. They chose the gesture uniquely mapped to the target to avoid confusion. *"For 'Stapler', we could perform the gesture of pressing it instead of holding it, because few other objects require pressing."* [P8]

From the questionnaires that participants filled in after the experiment, we count their ratings for the experience of using VirtualGrasp. The ratings were in five-point Likert scale. The participants showed very positive attitudes to retrieving objects with VirtualGrasp. They could easily discover the object-gesture mappings (AVG = 4.2, SD = 0.78), felt little fatigue through the interaction process (AVG = 4.4, SD = 0.70), felt

easy to recall the mappings (AVG = 4.5, SD = 0.53) and liked to use VirtualGrasp to retrieve objects in VR applications (AVG = 4.4, SD = 0.52).

DISCUSSION

Through three studies, we completed the end-to-end exploration of the grasping gesture design, from object list generation, gesture elicitation study, gesture classifier development to the usability study. Based on the results, we discuss the mapping strategies of how users mapped gestures to objects, the potential sweet spot applications, and the performance of VirtualGrasp with different sensing techniques.

Mapping Strategy

The core design of VirtualGrasp is the mapping metaphor of objects and gestures. The metaphor was to retrieve objects with the gestures to grasp their physical counterparts in reality. Under this metaphor, we probed what factors help users map a grasping gesture to the object. Through Study 1, we found that the usage, shape, and size played the most significant role. However, in Study 2, we found that sometimes one of these factors may dominate the mapping. For example, users mapped very similar gestures to the very small objects regardless of their shapes or usages. Besides, we found that the background and experience of the composer also influence the mapping strategy. Users who grasped the object every day would design different gestures from those who never used it.

Suitable Applications

We think VR games and teaching applications would benefit from VirtualGrasp. In some battle games, players frequently switch weapons or other game props. Using VirtualGrasp to retrieve them could enable players to focus on the ongoing task, enemies, and surroundings, without switching to the menu interface. In Study 3, users also reported that they would feel cool to retrieve weapons with VirtualGrasp. In a cooking teaching application, users could retrieve a bowl, spoon, knife, cup easily and pay main attention to the flow of the operations instead of searching for the tools. These tasks require users to intensively focus on the ongoing task, and VirtualGrasp could help them retrieve tools with few distractions.

Sensing Techniques

In our implementation, we use inertial sensors to track users' hands. However, there are other sensing options including data gloves [46], EMG sensors [32], depth cameras [3, 28], and vision-based techniques [4, 5]. We tested several vision-based techniques in our pilot study, and found both marker-based (Optitrack) and marker-less (LeapMotion) encountered the finger occlusion problem. To apply VirtualGrasp to these sensing techniques, the classification algorithm need to be designed [41] to solve the occlusion problem. For other sensors (EMG), the performance of using VirtualGrasp will rely on their tracking accuracy and we will test them in the future.

LIMITATION AND FUTURE WORK

In this paper, we focus on the design, implementation, and evaluation of VirtualGrasp, leaving several limitations and future work to be completed.

System Implementation

In our three studies, the participants were all right-handed. So our gesture recognition algorithm did not address the handedness issue, and left-handed users may be not able to use VirtualGrasp directly. However, in the future, we can adjust our algorithm to first recognize which hand of users is performing the gestures, and then apply different models to recognize right-handed, left-handed or both-handed gestures. In our implementation, to retrieve an object requires a dwell time of 0.5 seconds to trigger and then the classification results were returned in real time. In the future, we will optimize this speed and evaluate the efficiency of VirtualGrasp by comparing it to the current menus [39] and pointing techniques [37].

Object Set

In Study 1, the object list we generated aimed to represent graspable objects with different sizes, shapes. However, in Study 2, we noticed that some of the confusions of the grasping gestures were caused by the similar appearances of the objects, e.g. a phone and an interphone, a pen and a toothbrush; others were caused by the similar grasping gestures of too small objects, a flash drive and an eraser. In the future, we will test a larger set of objects with more objects that have similar shapes and sizes. Suggested by the results of Study 2, the classification accuracy might drop then, but there are potential solutions to alleviate this problem: 1) even two objects are of similar shape (e.g. a cue and a javelin), their usages might be different, which lead to different grasping gestures; 2) even the static grasping gestures of two objects look similar, e.g. broom and spear, the dynamic gestures are possibly still different, e.g. swiping the broom v.s.poking the spear. However, we think that it is not the best case to use VirtualGrasp to retrieve tools all with very similar appearances.

CONCLUSION

We propose VirtualGrasp, an object retrieval approach for VR applications. With VirtualGrasp, a user retrieves a virtual object by performing an in-air gesture which he/she uses to grasp or interact with its physical counterpart. Through three studies, we evaluated the consistency of object-gesture mappings across users, the expressivity of the grasping gestures and the performance of retrieving objects using this approach. Results have confirmed that users could reach high agreement on the mappings, that the object-gesture pairs could be accurately recognized by algorithms, and that users could discover the gestures by themselves and also enjoyed the experience. We also discuss the design implications, potential applications and limitations of this approach.

ACKNOWLEDGEMENT

This work is supported by the National Key Research and Development Plan under Grant No. 2016YFB1001200, the Natural Science Foundation of China under Grant No. 61672314 and No. 61572276, Tsinghua University Research Funding No. 20151080408, and also by Beijing Key Lab of Networked Multimedia.

REFERENCES

1. 2017a. Amazon product dimension data. Website. (2017). Retrieved March 28, 2017 from `https://www.amazon.com`.

2. 2017b. Destinations/SteamVR Environment. Website. (2017). Retrieved March 28, 2017 from https://developer.valvesoftware.com/wiki/Destinations/SteamVR_Environment.
3. 2017c. Kinect - Windows app development. Website. (2017). Retrieved March 28, 2017 from https://developer.microsoft.com/en-us/windows/kinect.
4. 2017d. Leap Motion. Website. (2017). Retrieved March 28, 2017 from https://www.leapmotion.com/.
5. 2017e. OptiTrack - Motion Capture Systems. Website. (2017). Retrieved March 28, 2017 from http://optitrack.com/.
6. 2017f. Rec Room. Website. (2017). Retrieved March 28, 2017 from http://store.steampowered.com/app/471710/Rec_Room/.
7. 2017g. Wikipedia website for fruits. Website. (2017). Retrieved March 28, 2017 from https://en.wikipedia.org/wiki/Watermelon.
8. Shaikh Shawon Arefin Shimon, Courtney Lutton, Zichun Xu, Sarah Morrison-Smith, Christina Boucher, and Jaime Ruiz. 2016. Exploring Non-touchscreen Gestures for Smartwatches. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3822–3833. DOI: http://dx.doi.org/10.1145/2858036.2858385
9. Ferran Argelaguet and Carlos Andujar. 2013. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics* 37, 3 (2013), 121–136.
10. Thomas Baudel and Michel Beaudouin-Lafon. 1993. Charade: Remote Control of Objects Using Free-hand Gestures. *Commun. ACM* 36, 7 (July 1993), 28–35. DOI: http://dx.doi.org/10.1145/159544.159562
11. Patrick Baudisch, Henning Pohl, Stefanie Reinicke, Emilia Wittmers, Patrick Lühne, Marius Knaust, Sven Köhler, Patrick Schmidt, and Christian Holz. 2013. Imaginary Reality Gaming: Ball Games Without a Ball. In *Proceedings of the 26th Annual ACM Symposium on User Interface Software and Technology (UIST '13)*. ACM, New York, NY, USA, 405–410. DOI: http://dx.doi.org/10.1145/2501988.2502012
12. Doug A. Bowman and Chadwick A. Wingrave. 2001a. Design and Evaluation of Menu Systems for Immersive Virtual Environments. In *Proceedings of the Virtual Reality 2001 Conference (VR'01) (VR '01)*. IEEE Computer Society, Washington, DC, USA, 149–. http://dl.acm.org/citation.cfm?id=580521.835855
13. Doug A Bowman and Chadwick A Wingrave. 2001b. Design and evaluation of menu systems for immersive virtual environments. In *Virtual Reality, 2001. Proceedings. IEEE*. IEEE, 149–156.
14. Andrew Bragdon and Hsu-Sheng Ko. 2011. Gesture Select:: Acquiring Remote Targets on Large Displays Without Pointing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 187–196. DOI: http://dx.doi.org/10.1145/1978942.1978970
15. Marcus Carter, Eduardo Velloso, John Downs, Abigail Sellen, Kenton O'Hara, and Frank Vetere. 2016. PathSync: Multi-User Gestural Interaction with Touchless Rhythmic Path Mimicry. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 3415–3427. DOI: http://dx.doi.org/10.1145/2858036.2858284
16. Nathan Cournia, John D. Smith, and Andrew T. Duchowski. 2003. Gaze- vs. Hand-based Pointing in Virtual Environments. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems (CHI EA '03)*. ACM, New York, NY, USA, 772–773. DOI: http://dx.doi.org/10.1145/765891.765982
17. Kaushik Das and Christoph W Borst. 2010. An evaluation of menu properties and pointing techniques in a projection-based VR environment. In *3D User Interfaces (3DUI), 2010 IEEE Symposium on*. IEEE, 47–50.
18. Geoffrey M Davis. 1995. Virtual reality game method and apparatus. (June 13 1995). US Patent 5,423,554.
19. Augusto Esteves, Eduardo Velloso, Andreas Bulling, and Hans Gellersen. 2015. Orbits: Gaze Interaction for Smart Watches Using Smooth Pursuit Eye Movements. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software & Technology (UIST '15)*. ACM, New York, NY, USA, 457–466. DOI: http://dx.doi.org/10.1145/2807442.2807499
20. Anthony G Gallagher, E Matt Ritter, Howard Champion, Gerald Higgins, Marvin P Fried, Gerald Moses, C Daniel Smith, and Richard M Satava. 2005. Virtual reality simulation for the operating room: proficiency-based training as a paradigm shift in surgical skills training. *Annals of surgery* 241, 2 (2005), 364.
21. Dominique Gerber and Dominique Bechmann. 2005. The spin menu: A menu system for virtual environments. In *Virtual Reality, 2005. Proceedings. VR 2005. IEEE*. IEEE, 271–272.
22. Tovi Grossman and Ravin Balakrishnan. 2006. The Design and Evaluation of Selection Techniques for 3D Volumetric Displays. In *Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology (UIST '06)*. ACM, New York, NY, USA, 3–12. DOI:http://dx.doi.org/10.1145/1166253.1166257
23. 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 (UIST '10)*. ACM, New York, NY, USA, 3–12. DOI: http://dx.doi.org/10.1145/1866029.1866033

24. Sean Gustafson, Christian Holz, and Patrick Baudisch. 2011. Imaginary Phone: Learning Imaginary Interfaces by Transferring Spatial Memory from a Familiar Device. In *Proceedings of the 24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 283–292. DOI: http://dx.doi.org/10.1145/2047196.2047233

25. Chris Harrison, Robert Xiao, Julia Schwarz, and Scott E. Hudson. 2014. TouchTools: Leveraging Familiarity and Skill with Physical Tools to Augment Touch Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2913–2916. DOI: http://dx.doi.org/10.1145/2556288.2557012

26. Maria Karam and others. 2005. A taxonomy of gestures in human computer interactions. (2005).

27. Hannes Kaufmann, Dieter Schmalstieg, and Michael Wagner. 2000. Construct3D: a virtual reality application for mathematics and geometry education. *Education and information technologies* 5, 4 (2000), 263–276.

28. David Kim, Otmar Hilliges, Shahram Izadi, Alex D. Butler, Jiawen Chen, Iason Oikonomidis, and Patrick Olivier. 2012. Digits: Freehand 3D Interactions Anywhere Using a Wrist-worn Gloveless Sensor. In *Proceedings of the 25th Annual ACM Symposium on User Interface Software and Technology (UIST '12)*. ACM, New York, NY, USA, 167–176. DOI: http://dx.doi.org/10.1145/2380116.2380139

29. Christian Kray, Daniel Nesbitt, John Dawson, and Michael Rohs. 2010. User-defined Gestures for Connecting Mobile Phones, Public Displays, and Tabletops. In *Proceedings of the 12th International Conference on Human Computer Interaction with Mobile Devices and Services (MobileHCI '10)*. ACM, New York, NY, USA, 239–248. DOI: http://dx.doi.org/10.1145/1851600.1851640

30. Arun Kulshreshth and Joseph J. LaViola, Jr. 2014. Exploring the Usefulness of Finger-based 3D Gesture Menu Selection. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 1093–1102. DOI: http://dx.doi.org/10.1145/2556288.2557122

31. Shahzad Malik, Abhishek Ranjan, and Ravin Balakrishnan. 2006. Interacting with Large Displays from a Distance with Vision-tracked Multi-finger Gestural Input. In *ACM SIGGRAPH 2006 Sketches (SIGGRAPH '06)*. ACM, New York, NY, USA, Article 5. DOI: http://dx.doi.org/10.1145/1179849.1179856

32. Jess McIntosh, Charlie McNeill, Mike Fraser, Frederic Kerber, Markus Löchtefeld, and Antonio Krüger. 2016. EMPress: Practical Hand Gesture Classification with Wrist-Mounted EMG and Pressure Sensing. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM, New York, NY, USA, 2332–2342. DOI: http://dx.doi.org/10.1145/2858036.2858093

33. Meredith Ringel Morris, Anqi Huang, Andreas Paepcke, and Terry Winograd. 2006. Cooperative Gestures: Multi-user Gestural Interactions for Co-located Groupware. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '06)*. ACM, New York, NY, USA, 1201–1210. DOI: http://dx.doi.org/10.1145/1124772.1124952

34. Miguel A. Nacenta, Yemliha Kamber, Yizhou Qiang, and Per Ola Kristensson. 2013. Memorability of Pre-designed and User-defined Gesture Sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '13)*. ACM, New York, NY, USA, 1099–1108. DOI: http://dx.doi.org/10.1145/2470654.2466142

35. Dennis A Nowak. 2009. *Sensorimotor control of grasping: physiology and pathophysiology*. Cambridge University Press.

36. Brandon Paulson, Danielle Cummings, and Tracy Hammond. 2011. Object Interaction Detection Using Hand Posture Cues in an Office Setting. *Int. J. Hum.-Comput. Stud.* 69, 1-2 (Jan. 2011), 19–29. DOI: http://dx.doi.org/10.1016/j.ijhcs.2010.09.003

37. Krzysztof Pietroszek, James R. Wallace, and Edward Lank. 2015. Tiltcasting: 3D Interaction on Large Displays Using a Mobile Device. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology (UIST '15)*. ACM, New York, NY, USA, 57–62. DOI:http://dx.doi.org/10.1145/2807442.2807471

38. Ivan Poupyrev, Mark Billinghurst, Suzanne Weghorst, and Tadao Ichikawa. 1996. The Go-go Interaction Technique: Non-linear Mapping for Direct Manipulation in VR. In *Proceedings of the 9th Annual ACM Symposium on User Interface Software and Technology (UIST '96)*. ACM, New York, NY, USA, 79–80. DOI: http://dx.doi.org/10.1145/237091.237102

39. Gang Ren and Eamonn O'Neill. 2013. 3D selection with freehand gesture. *Computers & Graphics* 37, 3 (2013), 101–120.

40. Jaime Ruiz, Yang Li, and Edward Lank. 2011. User-defined Motion Gestures for Mobile Interaction. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 197–206. DOI: http://dx.doi.org/10.1145/1978942.1978971

41. Toby Sharp, Cem Keskin, Duncan Robertson, Jonathan Taylor, Jamie Shotton, David Kim, Christoph Rhemann, Ido Leichter, Alon Vinnikov, Yichen Wei, Daniel Freedman, Pushmeet Kohli, Eyal Krupka, Andrew Fitzgibbon, and Shahram Izadi. 2015. Accurate, Robust, and Flexible Real-time Hand Tracking. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI '15)*. ACM, New York, NY, USA, 3633–3642. DOI: http://dx.doi.org/10.1145/2702123.2702179

42. Frank Steinicke, Timo Ropinski, and Klaus Hinrichs. 2006. Object selection in virtual environments using an improved virtual pointer metaphor. *Computer Vision and Graphics* (2006), 320–326.

43. Christian Steins, Sean Gustafson, Christian Holz, and Patrick Baudisch. 2013. Imaginary Devices: Gesture-based Interaction Mimicking Traditional Input Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 123–126. DOI: http://dx.doi.org/10.1145/2493190.2493208

44. Lode Vanacken, Tovi Grossman, and Karin Coninx. 2007. Exploring the effects of environment density and target visibility on object selection in 3D virtual environments. In *3D User Interfaces, 2007. 3DUI'07. IEEE Symposium on*. IEEE.

45. Radu-Daniel Vatavu. 2012. User-defined Gestures for Free-hand TV Control. In *Proceedings of the 10th European Conference on Interactive Tv and Video (EuroiTV '12)*. ACM, New York, NY, USA, 45–48. DOI: http://dx.doi.org/10.1145/2325616.2325626

46. Radu-Daniel Vatavu and Ionu Alexandru Zaii. 2013. Automatic Recognition of Object Size and Shape via User-dependent Measurements of the Grasping Hand. *Int. J. Hum.-Comput. Stud.* 71, 5 (May 2013), 590–607. DOI: http://dx.doi.org/10.1016/j.ijhcs.2013.01.002

47. Julie Wagner, Eric Lecolinet, and Ted Selker. 2014. Multi-finger Chords for Hand-held Tablets: Recognizable and Memorable. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2883–2892. DOI: http://dx.doi.org/10.1145/2556288.2556958

48. Jacob O. Wobbrock, Meredith Ringel Morris, and Andrew D. Wilson. 2009. User-defined Gestures for Surface Computing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 1083–1092. DOI: http://dx.doi.org/10.1145/1518701.1518866

49. Mike Wu and Ravin Balakrishnan. 2003. Multi-finger and Whole Hand Gestural Interaction Techniques for Multi-user Tabletop Displays. In *Proceedings of the 16th Annual ACM Symposium on User Interface Software and Technology (UIST '03)*. ACM, New York, NY, USA, 193–202. DOI:http://dx.doi.org/10.1145/964696.964718

50. Mike Wu, Chia Shen, Kathy Ryall, Clifton Forlines, and Ravin Balakrishnan. 2006. Gesture Registration, Relaxation, and Reuse for Multi-Point Direct-Touch Surfaces. In *Proceedings of the First IEEE International Workshop on Horizontal Interactive Human-Computer Systems (TABLETOP '06)*. IEEE Computer Society, Washington, DC, USA, 185–192. DOI: http://dx.doi.org/10.1109/TABLETOP.2006.19

51. Hans Peter Wyss, Roland Blach, and Matthias Bues. 2006. iSith-Intersection-based spatial interaction for two hands. In *3D User Interfaces, 2006. 3DUI 2006. IEEE Symposium on*. IEEE, 59–61.

52. Koji Yatani, Kurt Partridge, Marshall Bern, and Mark W. Newman. 2008. Escape: A Target Selection Technique Using Visually-cued Gestures. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '08)*. ACM, New York, NY, USA, 285–294. DOI:http://dx.doi.org/10.1145/1357054.1357104

53. Shumin Zhai, William Buxton, and Paul Milgram. 1994. The Silk Cursor: Investigating Transparency for 3D Target Acquisition. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '94)*. ACM, New York, NY, USA, 459–464. DOI: http://dx.doi.org/10.1145/191666.191822