IDS: The Intent Driven Selection Method for Natural User Interfaces

Frol Periverzov* Horea Ilieş†

Department of Mechanical Engineering
University of Connecticut

ABSTRACT

We present a new selection technique that facilitates the use of natural hand gestures for virtual object manipulation in 3D. Our method supports the use of 3D imaging techniques for tracking the user's body and therefore it does not require the use of any hand held devices that would restrict the manipulative capabilities of the user's hands. The key contribution of our work is the novel use of characteristic behavioral cues, which are representative for general goal directed movement, to infer the object targeted by the user during selection. The resulting technique enables us to select objects whose largest dimension is smaller than the sensing resolution of our system in spite of body tracking uncertainties and hand placement faults. Furthermore, by means of intention inference, our method automatically adapts to the user's subjective need for variable levels of tolerance to hand placement faults, jitter, or tracking noise.

Index Terms: H.5.2 [Information Interfaces and Presentation]: User Interfaces—User-centered design

1 INTRODUCTION

The selection task is one of the most common duties of our daily life. We select/choose our paths, our goals or objects of interest each time we decide to pursue a goal or to interact with this abstract object called "goal". Similarly, the selection of virtual objects is a common task of paramount importance for any virtual object manipulation process. The efficiency of the selection procedure directly impacts the performance of all other manipulation tasks such as virtual assembly. In this article we propose a new selection method that allows users to employ natural hand gestures in free space to select virtual objects. Our method facilitates the manipulation of virtual objects by means of natural hand gestures, and does not require the use of any hand held devices that would constrain the manipulative capabilities of the user's hands. In this endeavor we rely on 3D imaging techniques [20] to track the user's hands.

Several important challenges need to be overcome to achieve these goals. First, it is well-known that without physical support and/or haptic feedback, the users have difficulties in placing and holding their hands at the precise location required for selecting fine objects or details [4, 6, 13]. A common consequence of this problem is the penetration of the virtual objects located in the vicinity of the object with which the user intends to interact [12]. The second group of challenges is posed by the inherent noise and uncertainties that affect the 3D imaging methods, and the related stochastic tracking algorithms. *The tracking uncertainties* occur mainly for parts of the body that are affected by occlusion, imaging noise, low reflectivity, light glare, or body parts that cannot be distinguished from one another due to the perceived similarities caused by low imaging resolution or other factors.

*e-mail: frol.pv@gmail.com
†e-mail: ilies@engr.uconn.edu

IEEE Symposium on 3D User Interfaces 2015
23 - 24 March, Arles, France
978-1-4673-6886-5/15/$31.00 ©2015 IEEE

We address these challenges in a novel manner. The proposed technique identifies the objects that are targeted during the selection process by relying on a set of behavioral cues that have been documented in the neuropsychology literature for general goal directed actions. Such behavioral cues enable our method to tolerate hand placement and tracking faults. Some of the cues documented so far include facial cues [3], action efficiency [2, 8], action persistence [15, 10], effort invested, action duration [10], etc. By means of user studies we evaluate the relevance of two of the most promising cues in the context of the virtual object selection tasks. Specifically, our action efficiency cue estimates the effort required to select an object, while our action persistence behavioral cue estimates the level of perseverance with which the user tries to select a particular object. Our user studies show that by relying on the action efficiency cue our method affords the selection of objects that have their largest dimensions as small as 0.6 cm even when they are located in environments in which the distance to neighboring objects is approximately 0.1cm. Furthermore, embedding the action persistence cue along with the previous behavioral cue into our selection method enables users to select objects 45% faster and more efficiently than the case in which the action persistence cue is removed. The persistence cue allows our methods to detect the targeted objects during challenging selection tasks, when users show jittery or hesitant hand movements, in spite of the tracking noise that affects our system.

In this manuscript we demonstrate that by means of intention inference we can develop virtual object selection methods that enable users to efficiently select objects of dimensions smaller than the sensing resolution of their tracking system. Our seamless selection disambiguation method does not remove the environmental context during the selection procedure. Furthermore, our method automatically adapts to the user's subjective need for hand placement fault tolerance as described in section 3. Therefore, the work presented here advances the state of the art in 3DUIs towards more user-friendly or even person centered user interfaces by developing user adaptable interfaces driven by intention inference. Such developments can dramatically shorten the time required by a novice user to start performing efficient virtual object manipulations.

2 RELATED WORK

There are two main approaches used to select a virtual object: the virtual hand selection metaphor and selection by pointing. In the first case the selection is performed through distance evaluations between a virtual hand model and the surrounding objects, while the latter approach measures the proximity with respect to a ray that is defined implicitly by the user. For example, the ray direction can be provided by the line that joins two points on the user's body, such as the eye to hand tip direction, or it can be projected from a tracked device, such as a stylus. The virtual hand selection metaphor affords the use of natural gestures, while the virtual pointing approach can offer selection procedures that lower the arm fatigue at the expense of a less natural selection procedure [1]. Many of the published selection methods use as input devices hand held hardware, which constrain the manipulative capabilities of our natural hand gestures. Since we are interested in developing 3DUIs that offer the manipulation flexibility provided by our natural hand

gestures, we will mainly focus on virtual hand selection (VHS) approaches that do not make use of such hand held devices.

The VHS method is used in most 3DUIs in which the manipulation of virtual objects is controlled by simulating the physical interactions between a virtual hand model and the manipulated objects. These techniques usually use wearable input devices such as data gloves. In its most simple form the VHS method will select the objects that intersect the virtual hand model [9, 17]. The Go-Go technique [21] adopts a similar approach, but it allows the user to select objects outside the volume defined by the arm reach by elongating the virtual representation of the arm. In [16] the selection is activated once the virtual hand model intersects the object(s), and a pinch gesture is detected, while the 3D Bubble Cursor [26] method selects the object that is closest to the center of a selection sphere.

In [25] an abstract selection model is presented that aims at representing a large group of existing selection techniques. The model is composed of two main factors: (1) The relative position between some selection volume and the object that is targeted during selection, and (2) An abstract function of the history of the two factors. This model has a promising power of representation, but has not been tested yet. The work in [18] is concerned with identifying the movement phases of the users' hands during general selection tasks. This article offers a comparative analysis between the behavior shown by users while reaching to select objects in real environments and virtual environments.

The methods presented in [22] use as input the data offered by a Kinect camera, and do not require the use of any hand held devices. The selection is accomplished by using a selection cone whose apex is kept fixed while the center of the cone base can be moved in a vertical plane by the movement of the user's hand. The objects that are intersecting the cone can be selected, and a menu based selection disambiguation method similar to SQAD [13] is employed.

2.1 Selection Disambiguation

With the method mentioned above [22] the user can perform a hand pull gesture in order to display a 2D menu that lists the objects being intersected by the selection cone. Then the selection is accomplished by picking one of the listed items. The smallest objects that were selected using this method were spheres of 10cm diameter placed at a minimum distance of 30cm from all other objects. While such menu based disambiguation methods can be extremely accurate, they remove the environmental context from the selection procedure, and reduce the user's sense of presence in the virtual environment.

The Expand method [4] is addressing this problem by displaying the objects that intersect a 3D cursor in a grid pattern that overlays the image of the virtual environment. The selection is then completed by pointing a hand held controller towards the targeted object. The IntenSelect [5] method projects a selection cone from a hand held stylus, and the selection disambiguation is accomplished using a scoring function, which depends on the location of each object with respect to the cone axis and apex, previous scoring values and other tunable factors. In the Starfish [28] method the four closest objects to a 3D cursor are joined by a guiding surface. Once a target object is intersected by this surface, the user can press a button to lock the position of the guiding surface. Then, the 3D cursor is constrained to move inside this guiding surface. In this manner, the effort of positioning the 3D cursor with the high accuracy required by certain selection cases is significantly reduced. On the other hand, the steps involved in the selection process do not allow us to interact with virtual objects by means of free natural gestures. There are many other pointing techniques that have been proposed [7, 13], and most of them are reviewed in [1].

The method we propose here affords the use of natural hand gestures for selection, while using the hand wrist tracking method employed in [22]. Our user studies show that the smallest objects that can be repeatedly selected are spheres of 0.6 cm diameter located in an environment in which the distance to the neighboring objects is approximately 0.1 cm. This performance is achieved in spite of the fact that the depth sensing resolution of our current body tracing technique is 1 cm. We arrive at this result by developing a seamless selection disambiguation method that adapts to the user's need for hand placement and tracking fault tolerance as described in section 3.3.

3 Concepts and Implementation

3.1 General Concepts

The practical necessity for selecting objects that exhibit dimensions smaller than 1 cm becomes apparent when we consider manipulating vertices or edges located in a cluttered environment, or small geometric models of objects like bolts, nuts, chips, etc. Selecting such fine details proves to be surprisingly difficult in the context of the aforementioned hand placement imprecision and tracking uncertainties.

We overcome these issues by inferring the user's intent to select a particular object based on a set of behavioral cues that have been documented in the neuropsychology literature for general goal directed actions. Below we offer a conceptual description of the role played by each of the behavioral cues that we employ.

Using a metric for the efficiency of an action as an indicator for an intentional action is justified by *the principle of rational action* [8, 2]. This principle states that we, as rational beings, devise our actions such that we approach our goal in one of the most efficient manners, considering the constraints of the situation. It is therefore likely that the object which the user intends to select is among the objects that can be more easily or efficiently selected in the situation at hand.

The work in [15, 10] reveals that the quality of action persistence, or the fact that an action repeatedly ends in a similar state, represents significant evidence of an intentional action. Therefore the persistence shown by the user in approaching a particular object represents an important clue about the object that the user intends to select. Our hypothesis is that using an action persistence metric to infer the selection target will significantly improve the tolerance of our selection method to tracking and hand placement inaccuracies, especially during challenging selection tasks. We make this assumption based on the fact that challenging tasks require persistent and often repeated selection trials. It is known that a general hand reaching movement shows a ballistic phase [23] marked by a Gaussian-like wrist speed profile and a correction phase [18] that corresponds to the oscillating movement of the hand around the target position. The more challenging the selection task becomes for a particular user, the more prominent and longer in time the correction phase becomes. In all such cases, our action persistence behavioral cue will rapidly increase in value, and bias the inference towards the targeted object at an early stage, as described in section 3.3.1. In consequence, the correction phase will be significantly reduced in time. In section 4 we test this hypothesis by means of user studies.

It is important to consider that different people have different dexterity skills, and personal preferences regarding the manner in which they select and manipulate objects. In order to adapt to such personal differences, our selection method automatically adjusts the offered level of hand placement fault tolerance according to the subjective needs of the user. We evaluate the user's need for a certain level of fault tolerance by estimating the level of confidence shown by each user about the position in space of his/her hand. Observe that we naturally open our hands when we are uncertain about the position in space of our hands, or when we are preparing to grasp, and we move our finger tips closer to each other when we are ready to grasp, or when we are confident about the position of our hands. The correlation between the opening of our hand during a gen-

eral reach to grasp movement and the uncertainty we feel about the placement of our hand is supported by the principle of rational action [8, 2] described above. Namely, when we are uncertain about the position in space of our hands we open them widely in the attempt to increase our chances of reaching the targeted surface and therefore increase the efficiency of our hand reaching action. A similar observation can be made for the case in which we enter a dark room and attempt to explore the space using our hands. In consequence, we can use the opening of the user's hands along with motion cues that represent hesitant or oscillatory movements to estimate the level of confidence shown by each person about the position in space of his/her hand. The oscillations of the user's hands are captured by our action persistence behavioral cue which then controls the mechanism that compensates for hand placement and tracking faults. The same principle of rational action explains the fact that a person with lower hand control is instinctively opening their hand more in the attempt to perform a coarser object selection, or a power grip instead of a precision grip on the objects of interest. Such people also show hesitation or hand oscillations as soon as they face a selection task that they perceive to be difficult. Therefore by using the above behavioral cues, our system is able to estimate the user's subjective need for hand placement fault tolerance and adapt to it, as described in section 3.3.

3.2 System Setup

The current hardware setup for our virtual environment consists of a 3.95x1.672m Cyviz stereoscopic projective display that offers a rendering resolution of 2480x1050 pixels. In a plane approximately parallel to the screen a Kinect camera is placed to track the user's hand joints by using the algorithm presented in [24]. Finger motion tracking in 3D space was recently demonstrated by Softkinetic [11] using TOF imaging and in [19] using Kinect cameras and Leap Motion. Although these developments appear very promising, there is currently enough room for reliability improvements in their finger tracking capabilities. For this reason, we are developing the incipient stages of our interaction strategies using data gloves equipped with flex sensors for reading the flexure of the user's fingers. With this initial prototype our interface exhibits an effective workspace area of approximately $10m^2$. Before we feed the acquired tracking data into our intention inference algorithms, we pass it through an acceleration low pass filter to eliminate data indicating tracking faults or unnatural body motion. Once filtered, the tracking data drives a virtual hand model used to simulate the manipulation of virtual objects. We evaluate the collisions between virtual objects and simulate their interactions in a physically plausible manner using the PhysX engine, and perform scene rendering using UDK. The proper rendering of a virtual environment can significantly improve the user's spatial perception, and further mitigate the challenges induced by the lack of precision in hand positioning. It is important to note that human depth perception does not solely rely on the principles of stereo vision, but also on shading cues [27], motion cues as well as texture [14].

3.3 The Intent Driven Selection (IDS) Method

Our selection method offers hand placement fault tolerance according to the level of confidence shown by its users with respect to the position in space of their hands. Part of this tolerance is achieved by placing a proximity sphere around the simplified hand model of the user such that the fully extended fingers of the hand touch the interior surface of the sphere, as shown in figure 1. The proximity sphere is swept along the path described by the motion of the hand, and the objects that are intersected by it are considered to be candidate objects for selection.

The size of the proximity sphere is adjusted according to the users' level of confidence about the position of their hand. As the volume of this sphere corresponding to the fully extended fingers or hand placement uncertainty is considerably larger than the volume of the palm itself, the user can select objects with much lower hand positioning precision. Therefore the IDS method offers a higher degree of tolerance to hand positioning faults when the user shows such hand positioning uncertainty, and therefore needs a higher level of hand positioning tolerance. At the same time, once the users are confident about their hand position, the system offers them a lower hand positioning tolerance and concentrates the selection process on finer details by shrinking the proximity sphere as the hand closes (figure 1).

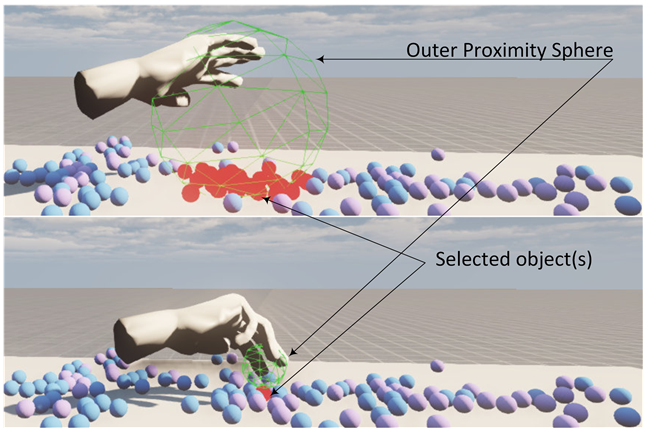

Figure 1: The adaptable proximity sphere

The selection of objects in cluttered environments can be controlled by placing a series of smaller proximity spheres inside of the outer proximity sphere as illustrated in figure 2. In this manner, the inner spheres of progressively smaller sizes are intersecting a subset of the objects intersected by the outer sphere as the user's hand approaches the target virtual object, which finally leads to a single object selection. The size of the inner proximity spheres automatically adapts to the user's intention in a similar manner to what was described for the outer proximity sphere. We will refer to this selection method as the IDS^1 method. During the selection process, the proximity spheres are invisible, and the selected objects change their color to red. Our tests show that the smallest objects that can be practically selected with this method are spheres of 4.5 cm diameter.

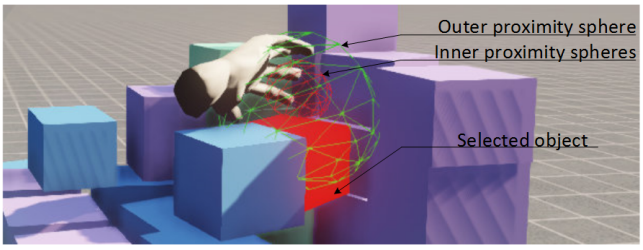

Figure 2: Progressive object selection in cluttered environments

In order to select finer details, we replace the above mentioned inner proximity spheres with the selection disambiguation mechanism described in the next section. By including the outer proximity sphere the resulting selection method, named IDS^2, incorporates and extends the adaptable hand placement fault tolerance and all other strengths offered by the IDS^1 method. While employing the IDS^2 method, the candidate objects for selection are highlighted using the glowing effect shown in figure 3. The object that is ultimately selected is marked red, and a green guiding beam joins the

center of grasp and the selected object. The beam is used to indicate the direction in which the users need to move in order to approach or depart the hand model from the object currently selected. We will use the term center of grasp to refer to the point located at an approximate distance of 4 cm from the center of the middle finger's middle phalange along the perpendicular to this phalange (figure 3). The location of this reference point has been established empirically.

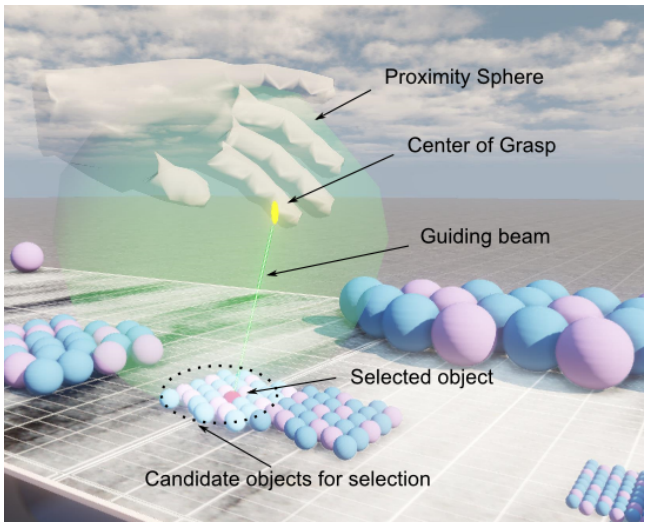

Figure 3: The graphical feedback generated during the selection process.

3.3.1 Selection Disambiguation

In order to be able to select in cluttered environments using natural hand gestures, we developed a seamless selection disambiguation mechanism that does not require the user to leave the current environmental context during the selection procedure. The proposed method selects the virtual object for which the following function is maximized:

$$iS(m,e) = t_l \cdot a_p(m,e) + a_{eff}(m,e) \quad (1)$$

$$a_{eff}(m,e) = \frac{1}{d_S(m,e)} \quad (2)$$

where iS - represents the strength of intent, m - the movement of the user's hand, e - the evaluated object, a_p - the action persistence and a_{eff} - the action efficiency , d_S - the distance to satisfaction, t_l - the tolerance/lock tuning factor.

The action persistence parameter captures the number of times in which the center of grasp lies in the Voronoi region of object 'e' during a specific time interval (figure 4). In other words, the a_p parameter estimates the number of user's attempts to approach object 'e'. The time window that we have used has a span of approximately 0.7 s while the position of the center of grasp is sampled approximately every 0.033 s. In order to evaluate the action persistence parameter, we use a proximity vector to store the identity of the object whose Voronoi region includes the center of grasp at the moment of sampling (see figure 4). The a_p parameter represents the number of times in which the object e was stored in the proximity vector during the past 0.7 seconds. The length of the time window that was used has been established empirically.

The d_S term represents the distance between the center of grasp and the surface of object 'e'. Therefore the action efficiency parameter will assume low values for distant objects that are difficult to select, and high values for objects that are close to the center of grasp. We use basic distance queries to evaluate the d_S parameter, as well as the membership of the center of grasp to Voronoi regions.

In this manner, for those users who show hand jitter, hesitation or lower hand control, the a_p behavioral cue will assume high values with respect to the object around which the user's virtual hand oscillates. As a result, our selection method identifies the target object at an early stage, and tolerates such hand placement or tracking faults. The tolerance is proportional to the value of the persistence behavioral cue as well as the size of the proximity sphere. In the case in which users show higher dexterity levels, our method will select the closest object to the users' hand.

The t_l factor is used to adjust the balance between the hand placement fault tolerance offered by our method, and the selection locking or sticking effect caused by large a_p values. Our experience shows that a good balance is achieved for $t_l = 2$.

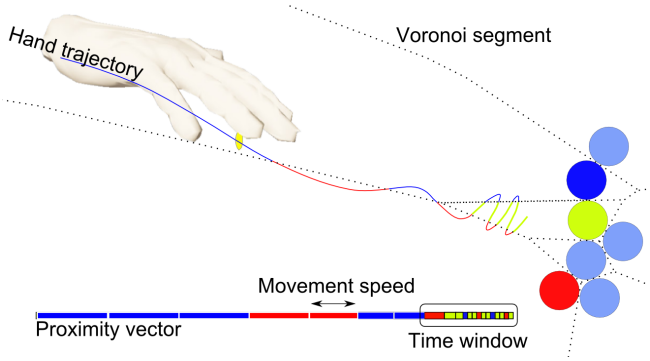

Figure 4: Evaluating the action persistence behavioral cue. The identity of each object is marked by its color, and the green disk represents the object that is targeted during selection.

4 Empirical Evaluation

In all studies described below the users stood approximately 2m in front of the projection screen described in section 3.2. A Kinect camera was placed in front of the screen to track the users movement. The study participants wore on their right hand a data glove equipped with 5 flex sensors that measure the approximate flexion of their fingers. Observe that the technique proposed in this paper is completely independent of the use of data gloves as long as the flexion of the fingers can be estimated. Different virtual environments were rendered on the screen for different tests, as discussed below.

4.1 Evaluating the Behavioral Cues

In what follows we test the main effects of the action efficiency and persistence behavioral cues on the performances of the proposed selection method. Specifically, we are interested in finding out the approximate size of the smallest object that can be practically selected when our disambiguation method is based solely on the action efficiency cue. Then we evaluate the hypothesis which states that by relying on the action persistence behavioral cue our method allows users to select their targets faster and more efficiently. Furthermore, we are investigating the influence of the users' number of selection trials using the IDS^2 method, and their previous experience with 3D virtual environments on the speed with which they manage to select their targets.

4.1.1 Test Population

Thirty participants were recruited to take part in this study. Their ages range from 20 to 52, with a median age 26, including 15 fe-

male participants and 3 left handed. Twelve have declared that they do not play video games or work with 3D CAD software packages and virtual environments, seven are sometimes using such 3D environments, and 11 use them frequently. The test lasted approximately 30 minutes and the participants were compensated 10$ for their time.

4.1.2 Procedure

Before taking part in the actual tests, the participants witnessed a brief (less than 20s) demonstration of the capabilities of the interface. Then, they were allowed to experiment by themselves with the elements of the interface for no more than 5 minutes. On the screen the virtual environment shown in figure 5 was displayed. The diameters of the spheres assigned as target objects for selection were: target one 4.5 cm, target two 2.25 cm, target three 1.12 cm, target four 0.6 cm.

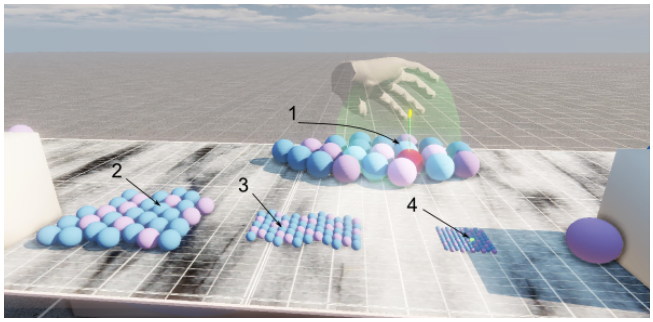

Figure 5: The selection test. Target number 4 is marked as the current selection target

In order to evaluate the efficiency of our selection method we considered the following performance parameters: 1) *Time efficiency*, which is the amount of time spent by the user while attempting to select designated objects, and 2) *Perceived effort*, which is the amount of effort spent by the user while performing the selection. To assess the effect of the action persistence behavioral cue on the efficiency of the IDS^2 method, we compared two versions of the IDS^2, namely with and without the a_p behavior cue. The most efficient selection method is considered to be the one that minimizes both performance parameters identified above.

To evaluate the above parameters, we asked the participants to select as fast as they can the objects marked as targets in figure 5. Only one object was designated for selection at a time. Selecting an object different than the designated object triggers a distinctive sound. In order to avoid potential confusion, a different sound is played once a new object is assigned for selection. At the same time, the target object starts blinking bright green (figure 5 target 4) until it becomes the subject of a *stable selection*. A selection is considered to be stable if the target object remains selected for a period of 2s, and no other object becomes selected during this period. The selected objects are colored red, while the candidate objects for selection are marked by the glowing effect shown in figure 3.

Once the user performs a stable selection, the system assigns a new target object. The time passed between the moment the target object is assigned and the moment the user completes a stable selection is recorded and used for measuring *the time efficiency parameter*. Following the procedure above, the system guides the user through all selection cases shown in figure 5. After iterating once through all cases, the user is notified that the selection method is switched to the other selection method. Then, the same procedure is followed while using the other selection method.

In order to minimize the influence of chance on the test outcomes, this process is repeated 10 times for each selection method and each user. As expected, the first selection trials are the slowest for each participant. To avoid biasing the data, the starting selection method is changed with each user. Therefore the tests are counterbalanced, and the 2 selection methods are evaluated in identical conditions. To avoid biasing the user's opinion, during the test we referred to the IDS^2 method that does not use the a_p behavior cue as *'the blue method'* while the other one as *'the green method'*. The color of the guiding beam was changed according to the names used. At the end of the test each user was asked to evaluate the following statement:

"The green selection method requires less effort than the blue selection method."

☐ strongly agree ☐ agree ☐ neutral
☐ disagree ☐ strongly disagree

The above Likert scale is used to evaluate the *perceived effort parameter*

We evaluate the size of the smallest spheres that can be practically selected when our selection disambiguation procedure relies only on the action efficiency cue by measuring the time spent by users while selecting, and the number of successful selection attempts on spheres of specific sizes.

4.1.3 Result Analysis

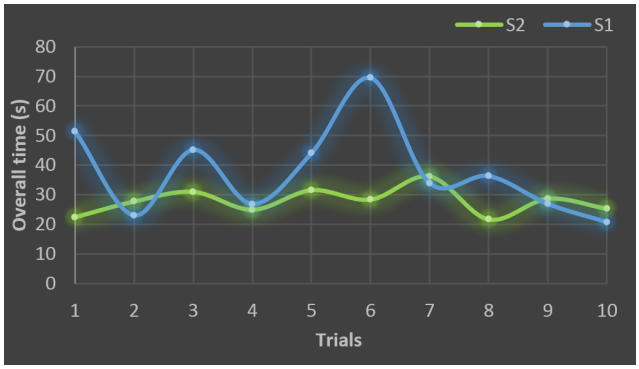

Figure 6: The evolution of the overall selection time of a typical user. S2 - the IDS^2 method employing the a_p behavioral cue, S1 - IDS^2 without a_p cue

The data shown in figure 6 suggests that the action persistence behavioral cue helps users achieve lower and less variable selection times. In order to obtain time efficiency parameters that are representative for the entire population, we average all data collected for each particular selection case. The results summarized in figure 7 indicate that by using the a_p behavioral cue the IDS^2 method becomes 5.4 % slower on target T1 (R = 2.25 cm) , 5.1 % faster on T2 (R = 1.12 cm) , 30.8 % faster on T3 (R = 0.56 cm) , 105.6 % faster on T4 (R = 0.28 cm) respectively 45.5 % faster overall.

We run repeated measures one way ANOVA tests to verify if the collected data provides significant evidences to support the above observations. The variances of the data collected for the 2 methods are stabilized by applying a natural logarithm transformation on the timing data.

The results show that, when augmented by the a_p behavioral cue, the IDS^2 method becomes significantly faster in terms of the time spent to select all 4 targets ($F_{1,29} = 83.7, p < 0.001, \eta^2 = 0.74$), as well as to select target 4 ($F_{1,29} = 129.9, p < 0.001, \eta^2 = 0.81$), and target 3 ($F_{1,29} = 13.6, p < 0.035, \eta^2 = 0.32$). On the other hand, the data does not show a significant difference between the methods

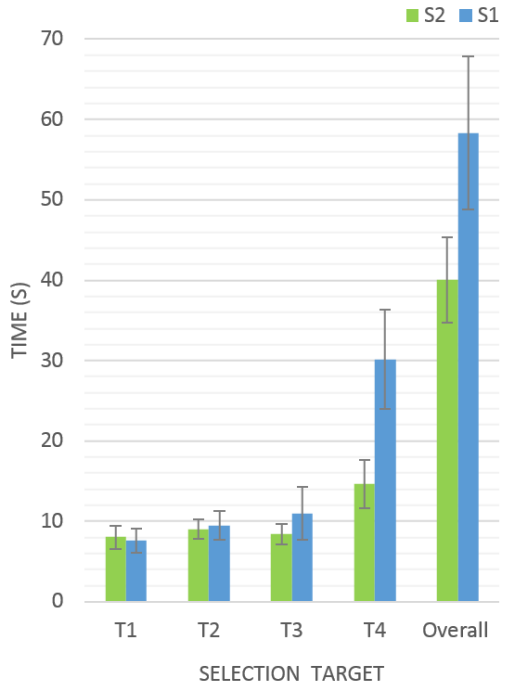

Figure 7: The selection time parameters averaged over the entire test population. S2 - the IDS^2 method employing the a_p behavioral cue, S1 - IDS^2 without a_p cue. The error bars represent the standard deviation of the average performance of each user.

when tested on target 2 ($F_{1,29} = 0.004, p > 0.1, \eta^2 = 0$) or target 1 ($F_{1,29} = 1.61, p > 0.1, \eta^2 = 0.05$). Also, as shown in figure 8, 23.3 % of the participants strongly agree that the 'green' selection method requires less effort than the 'blue' one, while 46.6 % agree, 13.3 % are neutral, and 16.6 % disagree. In consequence, we can conclude that by employing the action persistence behavioral cue our selection method allows users to select their targets faster and more efficiently, especially during difficult selection cases.

Out of the 300 selection trials that have been performed on target 4, 1% ended in abandon when the 'blue' selection method was used. We used a time measurement equal to 3 minutes for every task abandon. The mean and the standard deviation of the time spent by users during selection can be seen in figure 7. No abandon was encountered during the rest of the 2100 selection trials. Based on this data, we conclude that the smallest sphere that can be repeatedly selected while relying on the a_{eff} cue alone has a diameter of 0.6 cm. Furthermore during all 1200 selection trials no task abandon was encountered while using both behavioral cues in our selection disambiguation procedure.

Next, we look at the influence of users' experience with 3D virtual environments on their ability to perform quick selection tasks using the IDS^2 method. The tests that follow are performed with the complete IDS^2 method containing all behavioral cues presented in section 3. On the collected timing data we run a series of repeated measures ANOVA tests in which the declared user experience is treated as a 3 level factor. The results show that the users previous experience with 3D virtual environments does not significantly affect their performance in any of the 4 selection cases: T4 ($F_{2,27} = 2.42, p > 0.1, \eta^2 = 0.15$), T3 ($F_{2,27} = 1.08, p > 0.1, \eta^2 = 0.07$), T2 ($F_{2,27} = 1.75, p > 0.1, \eta^2 = 0.11$) and T1 ($F_{2,27} = 2.05, p > 0.1, \eta^2 = 0.13$).

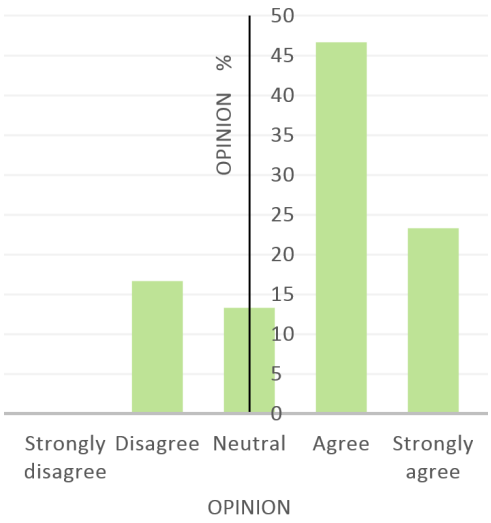

Figure 8: The user's perception of the selection effort reduction caused by involving the action persistence cue into the selection method.

Similarly, we test the influence of the number of selection trials on the users selection speed. This factor proves to be significant while tested on T4 ($F_{9,258} = 2.16, p < 0.05, \eta^2 = 0.07$), T3 ($F_{9,258} = 3.22, p < 0.001, \eta^2 = 0.1$), T2 ($F_{9,258} = 3.34, p < 0.001, \eta^2 = 0.1$) and T1 ($F_{9,258} = 7.85, p < 0.001, \eta^2 = 0.21$). Due to the fact that most of our users took a significantly longer time during their first selection trial than during the remaining 9 trials, we run the same ANOVA test without considering their first trial on each of the 4 targets. Interestingly, in this case the number of selection trials becomes insignificant when tested on T4 ($F_{8,229} = 1.04, p > 0.05, \eta^2 = 0.03$), as well as on T3 ($F_{8,229} = 1.71, p > 0.05, \eta^2 = 0.05$), and T2 ($F_{8,229} = 1.18, p > 0.05, \eta^2 = 0.03$), but not on T1 ($F_{8,229} = 6.02, p < 0.001, \eta^2 = 0.17$). These results show that after the first selection trial on each target the users stop learning how to use the IDS^2 method except in the case of target T1. This surprising exception could be explained by the fact that the tests on both selection methods start with T1. Therefore the users have at least the experience of selecting one T1 target before they attempt to select the other targets. The difference in the results obtained while considering the first selection trial, and the ones obtained while neglecting the first trial indicates that the IDS^2 method requires very little experience or training.

4.2 Comparing the IDS^1 Method With the VHS Method in Terms of Task Efficiency

Here we briefly perform a direct comparison between the efficiency shown by our IDS^1 method and the Virtual Hand Selection (VHS) method. As explained in section 3.3, the IDS^2 method incorporates and extends the strengths of the IDS^1 method and therefore the results obtained in this test represents an approximate lower bound of the capabilities of the IDS^2 method as well.

The test procedure is identical with what was presented in the previous test, except for the following aspects: eight volunteers took part in this study, including two female participants, and one left handed. Their ages range from 20 to 27, with a median age of 25. None of the participants had previous experience with manipulating virtual objects in 3D using natural gestures.

In order to evaluate the two performance parameters described in section 4.1.2, our participates were asked to select the targets shown

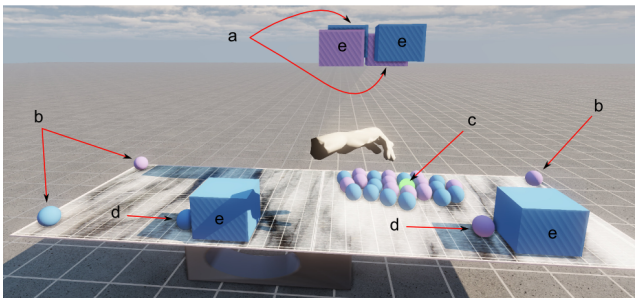

Figure 9: The selection cases: Selecting a) Large Occluded Objects, b) Under Tracking Noise Conditions, c) Small Objects in Cluttered Environment, d) Small Objects Close to Large Objects, e) Large Unoccluded Objects

in figure 9 five times with each selection method. The size of the virtual blocks (i.e., the *large* objects that we refer to below) used in our selection tests were $9 \times 10 \times 12$cm and the virtual spheres (the *small* objects) had a radius of 2.25cm. Each of the selection cases shown in figure 9 is designed to raise different selection challenges as we explain below:

- Case a): *selecting a large and mostly occluded object*. This task evaluates the selection efficiency in one of the common cases when users show a significantly increased hand placement imprecision. In such a case, the evaluation results will be dominated by the effects of *hand placement imprecision*. The motivation for this selection setup comes from the fact that users can observe only a small part of the target object and, therefore, they do not precisely know where the object boundaries are. Furthermore, because the target objects are occluded by large objects, the users' virtual hand will become partially occluded when the users approach their target. In consequence, participants will have difficulties in understanding the relative position between their hand and the target object, which decreases their ability to precisely position their hand model in the virtual space.

- Case b): *selecting a small object under tracking noise conditions*. As figure 9 shows, the target objects for this case are placed close to the lateral sides of the table model. The table top is positioned such that its side edges are in close proximity to the limits of the field of view (FOV) of the tracking camera. When the users try to get near these limits, parts of their body might leave the FOV of the camera or might get occluded by other body parts. In consequence, the image processing and tracking algorithms cannot collect sufficient information about the position of users body parts in order to produce reliable output. This fact translates into an increased frequency of tracking noise occurrence.

- Case c) *selecting a small object in a cluttered environment*.

- Case d) *selecting a small object positioned in close vicinity to large objects*.

- Case e) *selecting large and unoccluded objects*.

4.2.1 Result Analysis

During the test each participant performed with each selection method 5 selection trials on each of the 12 target objects marked in figure 9 to produce a total of 60 measurements for each participant and selection method. The timing data summarized in figure 10 suggests that on average the IDS^1 helps users select 76% faster in case a) in which users show hand placement imprecision, 26% faster while they are selecting under increased tracking noise conditions (case b), 15% slower when selecting small and cluttered targets (case c), 9% slower when selecting small objects that are in close proximity of large objects (case d) and 17% slower when selecting large and unoccluded objects (case e).

We use again one way repeated measures ANOVA tests to analyze the significance of the observations made above, as described in section 4.1.3. The results show that by using the IDS^1 method users select their targets significantly faster when facing the selection case a) ($F_{1,7} = 6.5, p < 0.05, \eta^2 = 0.48$) as well as in case b) ($F_{1,7} = 13.1, p < 0.01, \eta^2 = 0.65$). On the other hand the data does not show a significant difference between the methods in selection case c) ($F_{1,7} = 3, p > 0.1, \eta^2 = 0.3$) or d) ($F_{1,7} = 0.67, p > 0.1, \eta^2 = 0.08$). Surprisingly, the IDS^1 method turned to be 17% slower ($F_{1,7} = 26.1, p < 0.01, \eta^2 = 0.78$) than the VHS method in the case in which the users were asked to select large objects (case e) that do not require accurate hand placement or significant tolerance to tracking noise. This might be explained by the fact that while using the VHS method an object is selected once the hand model intersects an object. Because this selection case requires a lower level of control over the hand placement, and the user can easily see where the intersection takes place, the VHS method proves to be facile and fast. At the same time, unlike the IDS^2 method, the IDS^1 method does not show the extent of the proximity spheres. Therefore, without previous experience, the users cannot immediately tell where exactly is the intersection taking place, as is the case with the VHS or IDS^2 methods.

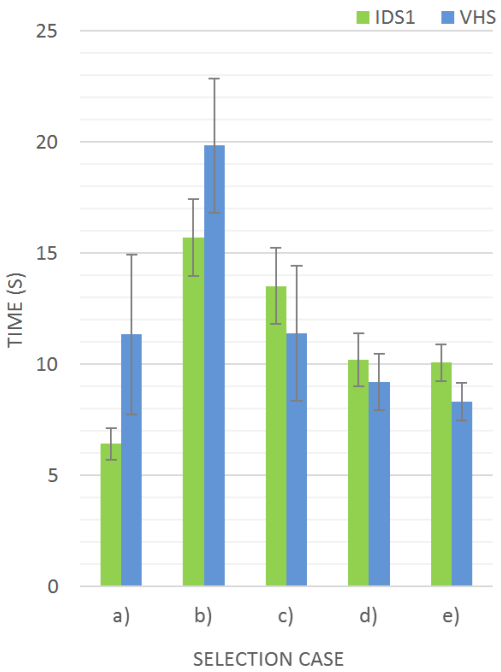

Figure 10: The selection time parameters averaged over the entire test population. The error bars represent the standard deviation of the average performance of each user.

While evaluating the perceived effort parameter 87.5% of the participants agreed that the IDS^1 method requires less effort for selection than the VHS method while 12.5% disagreed. These results indicate that the IDS^1 selection method allows users to select their targets faster and more efficiently than the VHS method, especially

during challenging selection cases.

5 Conclusions

In this manuscript we present a new virtual object selection method that facilitates the use of natural hand gestures to manipulate virtual objects in 3D. Our method does not rely on hand held devices or symbolic gestures and therefore, it does not restrict the manipulative capabilities of our natural hand gestures. Instead, this technique supports the use of 3D imaging methods for tracking the user's body, and compensates for the inherent tracking and hand placement faults.

When compared with the existent selection methods, our approach affords the use of natural hand gestures to select objects whose dimensions are smaller than the tracking resolution of the employed system. The proposed technique offers a seamless selection disambiguation mechanism, which does not require the user to leave the current manipulation context or use symbolic gestures and buttons.

We achieve these capabilities by identifying the objects that are targeted during the selection process based on a set of behavioral cues which have been documented into the neuropsychology literature. By means of user studies we have tested the relevance of 2 behavior cues with respect to the virtual object selection task. The results prove that the action persistence cue enables users to select objects 45% faster and more efficiently, especially during challenging selection tasks. At the same time the action efficiency behavior cue affords the selection of objects having their largest dimension as small as 0.6 cm even when these objects are located in environments in which the distance to neighboring objects is approximately 0.1 cm.

Furthermore, these behavioral cues enable us to estimate the user's need for hand placement fault tolerance during the selection process. In consequence, our method is capable of automatically adapting to the user's subjective need for various levels of hand placement and tracking fault tolerance.

Acknowledgements

This work was supported in part by the National Science Foundation grants CNS-0927105 CMMI-1200089.

References

[1] F. Argelaguet and C. Andujar. A survey of 3D object selection techniques for virtual environments. *Computers & Graphics*, 2012.

[2] E. Bonchek-Dokow and G. A. Kaminka. Towards computational models of intention detection and intention prediction. *Cognitive Systems Research*, 28:44–79, 2014.

[3] M. Carpenter, N. Akhtar, and M. Tomasello. Fourteen-through 18-month-old infants differentially imitate intentional and accidental actions. *Infant Behavior and Development*, 21(2):315–330, 1998.

[4] J. Cashion, C. Wingrave, and J. J. LaViola. Dense and dynamic 3d selection for game-based virtual environments. *Visualization and Computer Graphics, IEEE Transactions on*, 18(4):634–642, 2012.

[5] G. De Haan, M. Koutek, and F. Post. Intenselect: Using dynamic object rating for assisting 3D object selection. *IPT/EGVE*, pages 201–209, 2005.

[6] S. Frees. Context-driven interaction in immersive virtual environments. *Virtual reality*, 14(4):277–290, 2010.

[7] S. Frees, G. Kessler, and E. Kay. Prism interaction for enhancing control in immersive virtual environments. *ACM Transactions on Computer Human Interaction*, 14(1), 2007.

[8] G. Gergely and G. Csibra. Teleological reasoning in infancy: The naıve theory of rational action. *Trends in cognitive sciences*, 7(7):287–292, 2003.

[9] D. Holz, S. Ullrich, M. Wolter, T. Kuhlen, and J. Herder. Multi-contact grasp interaction for virtual environments. *Journal of Virtual Reality and Broadcasting*, 5(7):1860–2037, 2008.

[10] C.-T. Huang, C. Heyes, and T. Charman. Infants' behavioral reenactment of" failed attempts": exploring the roles of emulation learning, stimulus enhancement, and understanding of intentions. *Developmental psychology*, 38(5):840, 2002.

[11] S. iisu Product Datasheet V3.5.1. http://www.softkinetic.com, January 2013.

[12] J. Jacobs, M. Stengel, and B. Froehlich. A generalized god-object method for plausible finger-based interactions in virtual environments. In *3D User Interfaces (3DUI), 2012 IEEE Symposium on*, pages 43–51. IEEE, 2012.

[13] R. Kopper, F. Bacim, and D. Bowman. Rapid and accurate 3D selection by progressive refinement. In *3D User Interfaces (3DUI), 2011 IEEE Symposium on*, pages 67–74. IEEE, 2011.

[14] A. Loh and R. Hartley. Shape from non-homogeneous, non-stationary, anisotropic, perspective texture. In *Proc. of the BMVC*, pages 69–78. Citeseer, 2005.

[15] A. N. Meltzoff, A. Gopnik, and B. M. Repacholi. Toddlers' understanding of intentions, desires and emotions: Explorations of the dark ages. 1999.

[16] M. Moehring and B. Froehlich. Effective manipulation of virtual objects within arm's reach. In *Virtual Reality Conference (VR), 2011 IEEE*, pages 131–138. IEEE, 2011.

[17] M. Moehring and B. Froehlich. Natural interaction metaphors for functional validations of virtual car models. *Visualization and Computer Graphics, IEEE Transactions on*, 17(9):1195–1208, 2011.

[18] K. Nieuwenhuizen, L. Liu, R. van Liere, and J.-B. Martens. Insights from dividing 3d goal-directed movements into meaningful phases. *IEEE computer graphics and applications*, 29(6):44–53, 2009.

[19] I. Oikonomidis, N. Kyriazis, and A. Argyros. Efficient model-based 3D tracking of hand articulations using Kinect. *BMVC, Aug*, 2, 2011.

[20] F. Periverzov and H. Ilieş. 3D imaging for hand gesture recognition: Exploring the software-hardware interaction of current technologies. *3D Research*, 3(3):1–15, 2012.

[21] I. Poupyrev, M. Billinghurst, S. Weghorst, and T. Ichikawa. 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*, pages 79–80. ACM, 1996.

[22] G. Ren and E. O'Neill. 3d selection with freehand gesture. *Computers & Graphics*, 37(3):101–120, 2013.

[23] M. Santello, M. Flanders, and J. Soechting. Patterns of hand motion during grasping and the influence of sensory guidance. *The Journal of Neuroscience*, 22(4):1426–1435, 2002.

[24] J. Shotton, A. Fitzgibbon, M. Cook, T. Sharp, M. Finocchio, R. Moore, A. Kipman, and A. Blake. Real-time human pose recognition in parts from single depth images. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 1297–1304. IEEE, 2011.

[25] A. Steed. Towards a general model for selection in virtual environments. In *3D User Interfaces, 2006. 3DUI 2006. IEEE Symposium on*, pages 103–110. IEEE, 2006.

[26] L. Vanacken, T. Grossman, and K. Coninx. 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, 2007.

[27] G. Vogiatzis and C. Hernández. Practical 3D reconstruction based on photometric stereo. *Computer Vision*, pages 313–345, 2010.

[28] J. Wonner, J. Grosjean, A. Capobianco, and D. Bechmann. Starfish: a selection technique for dense virtual environments. In *Proceedings of the 18th ACM symposium on Virtual reality software and technology*, pages 101–104. ACM, 2012.