



Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube

Emmanuelle Chapoulie, Maud Marchal, Evanthia Dimara, Maria Roussou,
Jean-Christophe Lombardo, George Drettakis

► To cite this version:

Emmanuelle Chapoulie, Maud Marchal, Evanthia Dimara, Maria Roussou, Jean-Christophe Lombardo, et al.. Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube. [Technical Report] RT-0440, INRIA. 2013, pp.23. hal-00857534

HAL Id: hal-00857534

<https://inria.hal.science/hal-00857534>

Submitted on 3 Sep 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube

Emmanuelle Chapoulie, Maud Marchal, Evanthia Dimara, Maria Roussou, Jean-Christophe Lombardo, George Drettakis

**TECHNICAL
REPORT**

N° 440

September 2013

Project-Teams REVES, HYBRID
and DREAM and University of
Athens



Evaluation of Direct Manipulation using Finger Tracking for Complex Tasks in an Immersive Cube

Emmanuelle Chapoulie^{*†}, Maud Marchal[‡], Evanthia Dimara[§],
Maria Roussou^{§¶}, Jean-Christophe Lombardo^{||},
George Drettakis[†]

Project-Teams REVES, HYBRID and DREAM and University of Athens

Technical Report n° 440 — September 2013 — 23 pages

Abstract: We present a solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube). Rather than using a traditional wand device, users can manipulate objects with fingers of both hands in a close-to-natural manner for moderately complex, general purpose tasks. Our solution couples finger tracking with a real-time physics engine, combined with a heuristic approach for hand manipulation, which is robust to tracker noise and simulation instabilities. We performed an exploratory study to evaluate our interface, with tasks involving complex manipulations, such as balancing objects while walking in the cube. The user's finger-tracked manipulation was compared to manipulation with a 6-DOF wand (or flystick), as well as with carrying out the same task in the real world. Users were also asked to perform a free task, allowing us to observe their perceived level of presence in the scene. Our results show that our approach provides a feasible interface for immersive cube environments and is perceived by users as being closer to the real experience compared to the wand. However, the wand outperforms direct manipulation in terms of speed and precision. We conclude with a discussion of the results and implications for further research.

This project has been submitted to Virtual Reality journal in March 2013.

Key-words: virtual reality, direct manipulation, immersive cube, finger tracking

* emmanuelle.chapoulie@inria.fr

† REVES/INRIA Sophia-Antipolis

‡ HYBRID/INRIA Rennes, IRISA-INSA

§ University of Athens

¶ makebelieve

|| DREAM/INRIA Sophia-Antipolis

**RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Évaluation de la manipulation directe par finger tracking pour des tâches complexes dans un cube immersif

Résumé : Nous présentons une solution permettant d'interagir dans un système de réalité virtuelle de type cube immersif grâce au tracking des doigts. Plutôt que d'utiliser une manette traditionnelle, les utilisateurs peuvent manipuler les objets avec les doigts de façon quasi naturelle et ce pour des tâches générales modérément complexes. Notre solution associe le finger tracking à un moteur physique temps réel, ceci combiné à une approche heuristique de la manipulation manuelle qui est robuste aux signaux bruités des trackers et aux instabilités de simulation. Nous avons effectué une étude exploratoire afin d'évaluer notre interface. Elle comporte des tâches impliquant des manipulations complexes telles que tenir des objets en équilibre en même temps que l'on marche dans le cube. La manipulation avec finger tracking a été comparée à la manipulation avec une manette à 6 degrés de liberté, ainsi qu'à l'exécution de la tâche en condition réelle. Les utilisateurs ont également dû effectuer une tâche avec quasiment aucune instruction afin que nous puissions évaluer leur sensation de présence dans la scène. Nos résultats montrent que notre approche fournit une interface appropriée pour les cubes immersifs tout en étant perçue par les utilisateurs comme plus proche de l'expérience réelle qu'avec la manette. Cependant, la manette dépasse les capacités de la manipulation directe en termes de rapidité et de précision. Nous concluons par une discussion des résultats et des implications pour les prochains travaux de recherche.

Ce projet a été soumis au journal Virtual Reality en mars 2013.

Mots-clés : réalité virtuelle, manipulation directe, cube immersif, finger tracking

1 Introduction



Figure 1: A user in the four-sided cube holding a tray with two hands.

Interaction in immersive virtual reality systems (e.g., CAVes) has always been challenging, especially for novice users. In most systems, 6 degree-of-freedom (6DOF) devices such as “wands” or flysticks are used for navigation as well as selection and manipulation tasks. Such devices are well-established and can be very powerful; however, one goal of fully immersive systems is to enhance presence and immersion [23]. In such a context, these devices can potentially degrade the realism and naturalness of the virtual environment (VE). To avoid this shortcoming, we investigate the use of direct manipulation (DM) using finger tracking as close as possible to the manipulation used in the real world [19, 26, 14, 10].

With the advent of hand and finger tracking solutions, there has been recent interest in using DM and gestures in immersive systems to achieve specific tasks, such as automotive design [14, 19], focusing on carefully handling the physics of collisions between hands and virtual objects in specific tasks. However, little has been done to investigate the usability of DM in a fully immersive setting for relatively complex, general-purpose tasks. Moreover, the effect of DM on presence, or the similarity with real-world manipulations has not been sufficiently researched.

We present and evaluate a solution which incorporates DM for grasping and moving objects, using both one and two hands, based on finger tracking. Our system operates in a four-sided immersive cube (IC), i.e., a room with three rear projected walls and rear projected floor, providing a high sense of presence. To provide the most immersive and plausible experience, our solution uses a real-time physics simulator in addition to finger tracking, allowing close-to-natural manipulation of objects in the scene. We present an approach for DM which is robust to tracker noise and instabilities of the physics simulation.

The goals of our study are to evaluate: (a) whether DM is a feasible alternative to traditional IC interfaces such as a wand, (b) the effect of using DM on presence, and (c) the similarity to real world manipulation.

To demonstrate feasibility, we purposely choose a task that involves quite complex translations and rotations while the user walks in the cube, and at the same time balances objects on a tray. We compare DM to traditional wand-based interaction and, most importantly, compare to a real-world reproduction of the virtual task. We record both objective measurements (time to completion and errors) and subjective judgments through the use of a questionnaire.

Our contributions are twofold: a fully immersive solution which allows one and two-handed interaction in a four-sided retro-projected cube, and, more importantly, an exploratory study focused on the sense of immersion comparing our DM interface both with the traditional wand and with a real-world task. Our solution allows users to perform quite complex tasks, while benefiting from the realism and plausibility afforded by real-time physics and direct manipulation with fingers. In a nutshell, our study shows that two-handed DM enhances the sense of presence for some tasks, and users consider it more natural and closer to reality than the wand. Our objective measurements show that the wand and DM are in most cases equivalent in terms of speed and precision, but are slower than doing the task in the real world. The unconstrained user experience also shows several informal effects of enhanced presence due to DM, such as reflex reactions of participants trying to grasp dropped objects, or using two hands to adjust the position of a plate on a table.

The conclusions of our study provide insight into the usability of DM in a fully immersive four-sided cube, clearly demonstrating its utility and indicating some advantages over traditional 6DOF interfaces, in terms of immersion, presence, and similarity to the real world.

2 Previous Work

Gesture-based interaction has received significant interest in virtual or augmented reality research [24, 20, 6, 7]. A thorough review of natural gestures for virtual reality (VR) [5] underlines the many positive features of natural, but also discusses the utility of “hyper-natural” interfaces. In a previous survey of 3D user interfaces (UI) [4], natural gestures are mentioned as one of the important future directions for 3D UIs. In many cases, gestures are used to define a vocabulary or language [6] even if the number of gestures is often limited [24, 20, 7]. This can create an overhead for the user who must remember the meaning of each gesture. An early solution providing more natural interfaces involved the use of multimodal interactions, combining depictive gestures with speech [15, 17, 16]. We are more interested in the case of DM, i.e., interacting with the environment in a natural manner with the users hands [19, 26, 14, 10].

The addition of physics simulation to DM provides truly intuitive interaction with the objects in the scene. Such approaches for interface design have received much interest in recent years, often linked with tabletop systems [1, 27]. The use of physics simulation to provide natural feedback in the environment has also been of great interest in VR research. One remarkable early result was that of [8], which demonstrated the use of a fast physics solver and hand-based interaction, in the context of a workbench environment. The physics solver presented was one of the first providing sufficiently fast simulation to allow realistic interaction. A major difficulty is how to handle objects controlled by the users hands (often called “God objects”) with respect to the simulation of the rest of the environment, i.e., correctly providing external forces from the hands.

Over the last ten years both physics simulation solutions and gesture-based input hardware have progressed immensely, providing the ability for much more accurate simulation. Much previous work concentrates on the more technical aspects of gesture recognition and its integration with physics, and often includes the calculation of forces for haptic force-feedback systems [21]. Initial work, before physics simulation became widely available, focused on the definition of ap-

appropriate gestures for certain kinds of operations, and notably grasping [25]. Experiments with force-feedback systems allowed the simulation of several quite complex manipulations [11]. A spring model coupled with a commercially-available physics simulator was used in [3] to simulate various kinds of grasping operations with objects of varying complexity, and to avoid hand-object interpenetrations. A simpler approach was proposed in [12], where grasping is simulated without complex physics. Grasping and interpenetration were also the focus of the work in [22].

In [19], DM was used based on grasping heuristics and constraint-based solutions for the specific case of a car interior. A quantitative comparison to a real-world car interior (mirror, door etc.) was performed. More recent work has concentrated on developing appropriate soft models [13] and efficient solvers to avoid interpenetration of God objects and other objects in the scene [14], mainly in the context of automotive project review. The latter uses a setup similar to ours, with the same finger tracking system; however a three-sided display system is used, with the floor projection from above. Usability and presence were not studied in this work.

The use of consumer-level depth sensors also opens numerous possibilities for natural interaction with collocation of virtual hands and a virtual scene, albeit in a limited workspace subject to various system constraints e.g., the need for a split screen setup and the limited range of the depth sensor [10]. Our 4-sided IC also allows hand collocation and interaction, but does involve a number of hard constraints related to the geometry of the IC and the design of the tracking system. In particular, the three surrounding walls restrict the possible positions for the trackers, and result in relatively large regions of “shadow” for the tracking system.

Our focus will be on presence, usability and user satisfaction when using DM in fully immersive environments. Some recent work exists for example on the effect of using physics on task learning [2] or the use of natural interaction for video games [18]. [9] proposed to evaluate recognition methods through classification to avoid the calibration of gesture-based input devices. However, the study of usability for DM in general tasks, and in a fully IC-like environment with walking users has not received much attention.



Figure 2: Left: Visual representation of the palm and three fingertips in the VE (cubes are shown away from the fingers for clarity of illustration). Right: Wand selection used for comparison.

Our goal is to compare DM to the use of a wand and to real-world manipulations, in the challenging context of full immersion, which allows a close-to-natural interaction with the environment.

3 A Heuristic Approach for Direct Manipulation with Physics

There are several difficulties in developing a DM interface in an IC-like environment. In contrast to systems with a restricted workspace, in which the user is sitting [22, 10], we target a room-sized environment, and allow the user to walk around the scene while manipulating objects. There are three main difficulties we discuss below: finger tracking, dynamic constraints between objects and the physics simulation.

First, finger tracking in the IC is challenging. Even though we use a high-end finger tracking system (Sec. 4.2), it is prone to noise and interruptions in the signal for the fingers. In addition, a freely walking and moving user can often be in, or close to, “shadow” regions of trackers where the signal is deteriorated. This is due to occlusion from the user himself, and tracking calibration which is also hindered by the enclosing walls and the user occlusion. Tracking “shadow” regions thus occur in the zone 10-20cm away from the screen. To overcome these issues, we designed our tasks to avoid activities where the user’s body blocks the tracking cameras or involves objects in the “shadow” regions. We detail our treatment of finger tracking and tracker signal filtering in Sec. 3.1.

Second, given our goal of assessing feasibility of DM, we ask users to perform relatively complex tasks: grasping and translating objects, including balancing objects one on top of the other while walking. In previous work, users often manipulate a single dynamic object per hand, with simple constraints, e.g., collision detection with static objects. In contrast, our tasks involve several dynamic objects, and multiple indirect constraints from physics-based interactions between objects. This requires the robust tracking and grasping solutions we developed.

Third, our tasks require the use of a fast physics engine; we use *Bullet*¹ (see Sec. 3.4 for details), which is fast enough to handle quite complex scenes (e.g., our user experience scene in Fig. 14²) in real time. *Bullet* uses an impact-based simulation approach. As a result, objects tend to bounce between the fingers instead of being grasped, and the contacts are unstable. Many recent solutions improve physics simulation, most notably to avoid inter-penetration of the hand and other objects in the scene (e.g., [14, 22]), while other approaches [25, 12] avoid the need for complex – and often expensive – precise simulation using specific algorithms. In contrast, we use a fast but simple physics simulation, combining it with an approach for basic DM such as grasping and releasing, and a finite-state machine to handle sequences of manipulations (e.g., moving from one-hand to two-hand grasping etc.).

In addition, the physics simulation requires careful synchronization with the displays. The simulation is run on one machine, the transformation information of all dynamic objects is propagated to all slaves at each frame. We next discuss the solutions we adopted to address these problems.

3.1 Finger Tracking and Signal Filtering

The user is presented with a representation of the palm, thumb, index and middle fingertips in the form of small white cubes, providing visual feedback of hand position and orientation (see Fig. 2(left)). We used cubes for efficiency; they are mostly hidden by physical fingers, and thus did not interfere with interaction. In the following, only the thumb and index of each hand are used; we call these *active fingers*; the middle finger is ignored in our current implementation. Two *active fingers* proved to be largely sufficient to provide a natural-feeling DM interface, as seen in the evaluation.

¹<http://www.bulletphysics.org>

²Due to the low light condition in the IC, we took photos with a long exposure which results in some blur in the photographs.

When grasping with one hand, the *main contacts* are the actual contacts of the thumb and index with the object. When grasping with two hands, there is one *main contact* per hand. This is the actual contact between the finger and the object if only one finger is applied, or a “mean contact” if two fingers are applied with this hand. The “mean contact” is set at the midpoint between the two actual finger contacts and is useful for the grasp/release finite state machine described next.

The finger tracking signal in the IC lacks precision involving noise (“trembling”) and loss of signal (“jumps”). We apply a Kalman filter to each finger in time, and track the variance of the signal over a sliding window. If variance is above a threshold we test if there is a plateau in the signal, in which case we identify this as a loss of signal and do not apply the motion to the object.

3.2 Grasp/Release Heuristics

We directly use the physics engine to simulate the manipulations in the environment. Cubic objects are attached to the fingertips and the top of the palm. These are “kinematic objects” in *Bullet* terminology, and can apply forces to the other dynamic objects being simulated. Kinematic objects are not affected by other objects. For each object, we track contact events with the fingers provided by the physics simulation. Once two kinematic contacts on a given object are identified we mark the object as “grasped”, until the contacts are released. We remove grasped objects from the physics simulation and enforce the transformation and speed of the trackers so that the simulation of objects in contact with the object selected is still correct.

When grasping, we store specific data (e.g. position and orientation of the selected object), called *grasping data*, which is reset each time the number of contacts on the object changes. This data is used to compute the transformations of the selected object, depending on the fingers movements.

If we simply use the information from the tracker and the physics engine, users “drop” objects very quickly, or objects may move in an unstable manner. We make the selection more robust by marking an object as “released” only when the distance between the contacts is greater than 10% of the distance stored in the *grasping data*. It is important to understand that contacts are determined by the physics engine in a natural manner when treated with our approach; they are not “pinch”-like gestures as the selection is determined by the actual contacts and distance between contacts and not by the movements (see Fig. 4). If a hand participates in holding an object by applying a single finger, it releases as soon as the distance between the two main contacts is 10% longer than that stored in the *grasping data*. We chose these thresholds by pilot trial-and-error tests in several different settings. While releasing, the object being handled follows the translation of the middle point between the two *main contacts*. Accurately placing objects thus requires a short learning period (see the video and Fig. 3 for an example). However, users did not complain about this limitation.

The threshold for release is the main restriction of our approach in terms of “naturalness”. If a user grasps a “thick” object with one hand, it is not possible for her to directly release it in the current implementation if she cannot open her hand by at least 10% of the current gap. To release the selected object, the user would need to switch from a one-handed grasp to a two-handed grasp to reset the *grasping data* and then release the object. This scenario did not occur in our tasks.

As mentioned before, selected objects are not handled by the physics simulation. The translation applied is that of the midpoint between the main contacts, and the rotation applied is that of the vector defined by the main contacts. When released, the objects are re-inserted into the physics simulation.

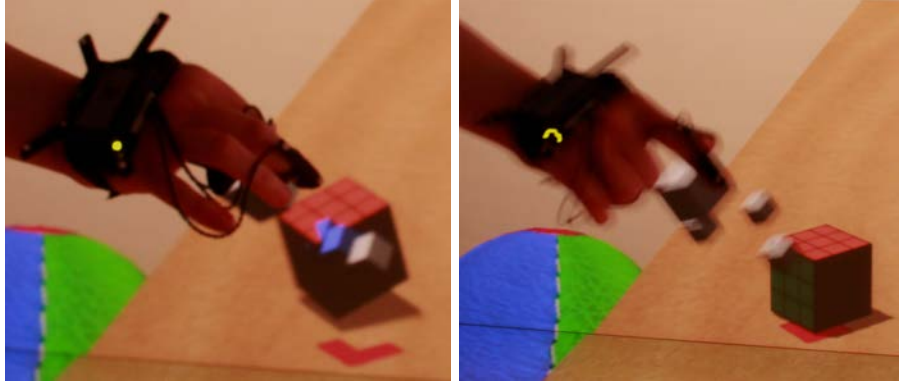


Figure 3: Accurately placing a cube in a corner.



Figure 4: A user spreads her arms to release a large object without the need to further open her hands.

3.3 Finite State Machine for Object Manipulation

When performing complex tasks, users naturally grasp and release objects, and often mix one and two-handed manipulations. To treat such transitions, we design a finite state machine approach. We next describe the different possible transitions for each manipulation.

Grasp. An object is grasped when the user applies at least two fingers on it, with only one hand or both hands (see Fig. 6, left). The user can switch from a one-handed grasp to a two-handed grasp by touching the object with one or both active fingers of the second hand.

There is no alignment test on contacts, so users could lift an object with two contacts on the same side of a cube. However, this would be unnatural, and no user attempted this gesture in our tests.

Release. An object is released when there is at most one finger in contact with the object. For a grasping hand to release the object, the user simply has to open her fingers (see Fig. 6, right). If one or both hands are grasping just by applying one of their active fingers, the user simply has to remove these hand(s) from the object. If both hands are grasping, the user can switch to a one-handed grasp by opening one hand. Finally, to completely release an object grasped with

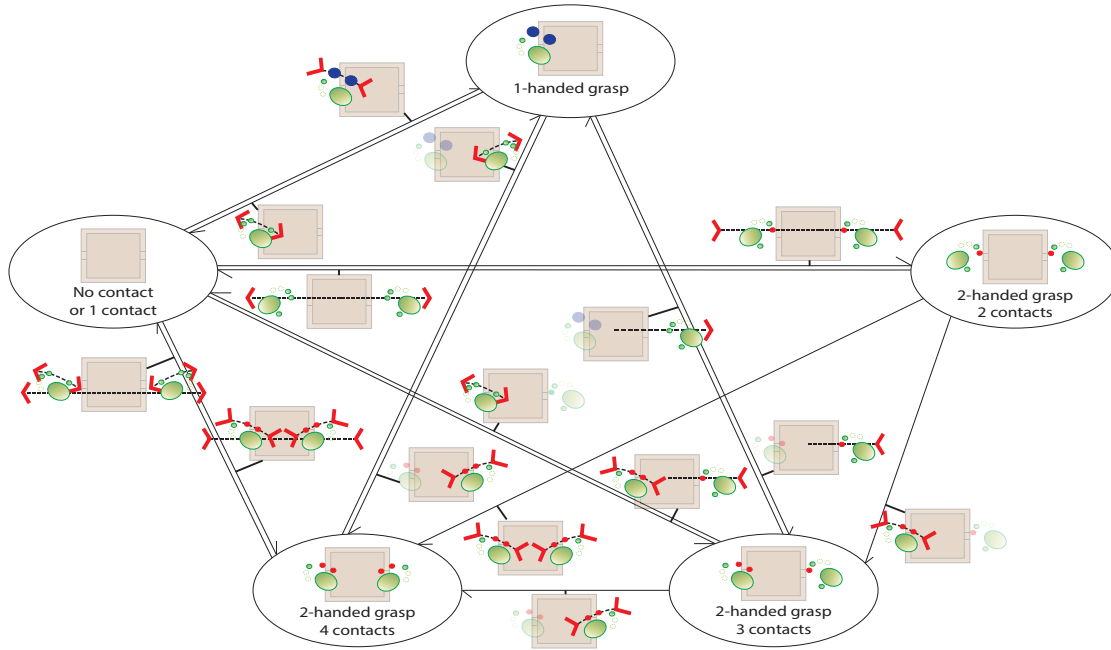


Figure 5: Finite state machine graph detailing transitions between grasping and releasing states. A representation of the hand is shown in green; dots represent fingers. When only one active finger is touching a dynamic object, the corresponding dot becomes plain green. When two active fingers of the same hand are grasping an object, the corresponding dots turn blue. When the object is grasped with both hands, the dots in contact with the object become red.

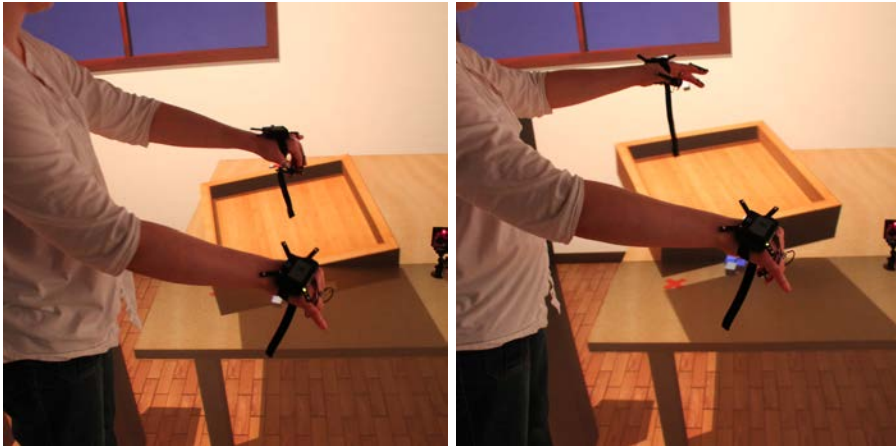


Figure 6: Left: two-handed grasp. Right: two-handed release.

two hands, the user must spread his arms to ensure that the fingers are no longer in contact with the object.

The full list of possible transitions is provided in Fig. 5.

Translate. The grasped object is translated when the fingers in contact translate.

Rotate. The grasped object is rotated as soon as the fingers in contact rotate. We currently limit the rotation with two hands to simplify the implementation, avoiding problems with the combined effect of tracker noise on each hand: the user can rotate one hand with respect to the other (e.g., tipping out balls from a tray).

The user can switch from a one-handed grasp to a two-handed grasp and vice versa; two one-handed grasps can be used at the same time to manipulate two distinct objects (Fig. 7, right).

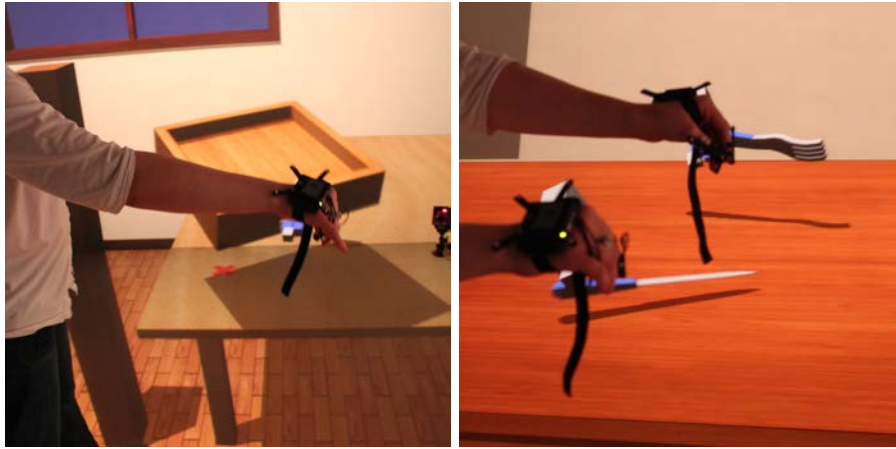


Figure 7: Left: one-handed grasp. Right: two hands grasping two objects.

3.4 Implementation

Color coding is applied to the representations of fingertips. By default, the cubes are white. When only one active finger is touching a dynamic object, the corresponding cube becomes green. When two active fingers of the same hand grasp an object, the corresponding cubes turn blue. When the object is grasped with both hands, the cubes in contact with the object become red.

We use our in-house VR software system which is based on OpenSceneGraph³. We adapted the Bullet Physics open source physics engine to communicate the data needed by our grasping/release approach and our finite state machine through extra object properties, as we enforce the transformations of the handled objects coherently with those controlled by the physics simulation. We calibrate the interocular distance, as well as Y-Z positions of the eyes for each participant at the beginning of each experiment. This provides a much better immersive experience in the IC by improving the projection from the user's point of view (Sec. 4.2). This is also more comfortable for the user as it is expected to reduce the risk of cybersickness.

³<http://www.openscenegraph.org/>

4 Exploratory User Study

Our goal is to evaluate whether DM is a feasible alternative to traditional interaction interfaces such as wands, whether it affects presence, and how similar it is to real world manipulations. To provide a meaningful evaluation of feasibility, we purposely choose a complex task, which involves quite complex translations and rotations, while walking and balancing objects on a tray (see Sec. 4.3 for details). We do this by comparing DM to a traditional wand-based interaction and by comparing both the wand and DM to a real-world task: the virtual world used is a replica of a real space in which the same tasks are performed (Fig. 12). Our hypotheses are : a) using the wand will be more precise and faster than finger-tracking based manipulation and 2) using finger-tracking based DM will be more natural and will provide a higher sense of presence.

We record both objective measurements, namely time to complete a task and precision or errors, and subjective judgments based on a questionnaire.

4.1 Population

We ran the experiment with 18 participants, 10 men and 8 women aged between 24 and 59 years old, (mean age 32.5 years, standard deviation 10.7 years). Most had no experience with Virtual Reality (13 out of 18); 5 had experienced VR demonstrations before. Four participants had previously used a wand, including three who had previously manipulated virtual objects.

4.2 Experimental Apparatus

We use a four-sided IC (the Barco iSpace⁴, comprising three walls and a floor), which has four retro-projected “black” screens. The front and side screens are 3.2m wide x 2.4m high, and the floor is 3.2m wide x 2.4m long. Stereo is provided using Infitec technology, and for tracking we use an ART⁵ infrared optical system, with eight cameras (see Fig. 8 left). We track the head with a frame mounted on glasses. The tracked devices provided are the wand and the finger-tracking system of ART to track the palm, thumb, index and middle finger of each hand (see Fig. 8 right). The finger-tracking system tracks the tips of the fingers giving the relative positions of these frames with respect to a 6DOF active tracker which is on the back of the hand.

Normally, six cameras are sufficient for head tracking. However, to allow better quality finger tracking and reduce “shadow” regions, two additional cameras were added. As these cameras must be placed in the bottom front corners of the IC and the cables connecting them to the system are visible in certain areas of the side walls, this can affect the user’s immersive experience. We thus designed our scenarios to avoid eye gaze on these areas as much as possible, and we also designed our models to hide the cables and the cameras by adding very dark baseboards to the room.

Wand Interaction For the wand, we use a standard virtual ray emanating from the wand. To grasp an object with the wand, the user points the wand towards it and presses a trigger button. Small blue spheres are displayed at contact points between the ray and the object when grasping (Fig. 2(right)). The trigger is kept pressed during manipulation. To release an object with the wand, the user simply has to release the trigger.

We implemented the ray to be thin to limit occlusion, and to be long enough (1m) allowing sufficient coverage of the 3.2mx3.2mx2.4m IC.

⁴<http://www.barco.com/en/products-solutions/visual-display-systems/3d-video-walls/multi-walled-stereoscopic-environment.aspx>

⁵<http://www.ar-tracking.com>

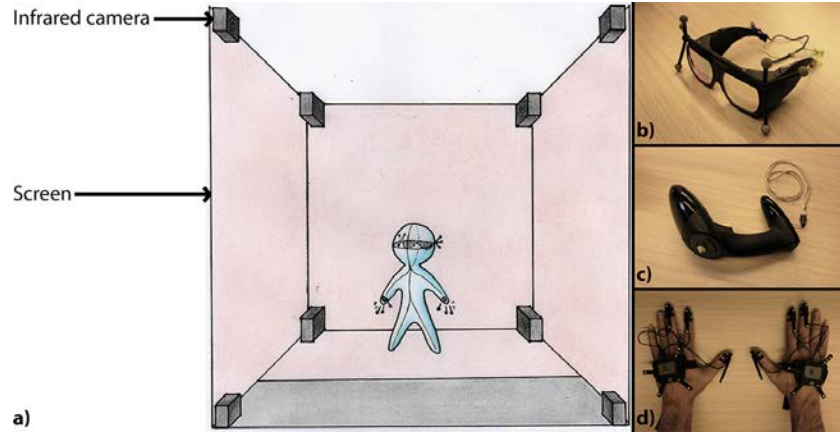


Figure 8: Experimental setup. The four-sided IC is represented in a) with its eight infrared cameras. On the right are photos of the tracked input devices used: b) Infitec glasses with frame, c) ART flystick and d) ART finger-tracking gloves.

4.3 Experimental Procedures and Environments

The experiment lasted 1h30 on average. It consists of a calibration step, a training session, a usability test, a “free form” user experience and questionnaire completion. The usability evaluation has three conditions: using the wand, DM, and the real world condition. Hence, we had six groups of three participants to test all the possible orders of conditions, which were randomized. For DM and real, the users always performed the tasks first with two hands, and then with only one hand. We decided not to vary this order within the conditions using hands to permit the user to learn progressively. Specifically, when handling larger objects such as a tray, using two hands provides better control and balance overall. In contrast, using one hand involves grasping the border of the virtual tray; slight motion of one finger with respect to the other can result in a large motion of the tray, making it harder to control. In the design of all tasks, we tried to minimize the cases of the hand incorrectly occluding virtual objects; evidently this is not always possible, but we did not receive any negative feedback about this from participants.

The training session and the free-form user experience session are not performed in the real condition, but the order of the other conditions is the same as in the usability task. Concerning the DM, training session and the usability task are performed with two hands and one hand separately whereas, in the user experience, the user can freely manipulate objects with one or both hands. The best way to appreciate these experimental procedures is to watch the accompanying video.

Calibration step:

The finger tracking devices are calibrated for each user so that we have more reliable signals. Both devices are calibrated separately to avoid interference, using the procedure defined by the manufacturer.

We also calibrate the position of the eyes. We performed a pilot study where we simply measured and set the interocular distance: this proved to be insufficient since the environment still displayed a “swimming” effect. To overcome this, we also adjust the other two coordinates of the eye positions (see Fig. 9). This is done using a simple scene, i.e. a floor and a stool (see

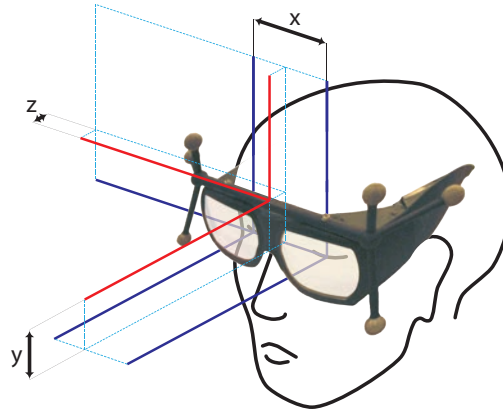


Figure 9: The position of the user's eyes is calibrated in the three dimensions.

Fig. 10, left). The experimenter progressively modifies the coordinates of the eye positions until the cubes representing the fingers are closer to the fingertips and the user feels comfortable with the projection. A subsequent step consists in modifying values to minimize perceived movement of static objects when moving in the IC.



Figure 10: Left: Calibration Scene. Right: Training scene

Training Session:

The goal of the training session is to familiarize the participant with the interaction techniques. The experimenter first explains the color coding of the cubes and how to use the techniques to grasp, release and move the objects.

To maximize immersion, the virtual scene consists of a closed room exactly the size of the actual IC, with two tables, and three cardboard posts in between (see Fig. 10, right). On the left table, the user is provided with a colored cube and a tray containing two balls. There are red crosses marked on the tables: one under the cube on the left table, and two on the right table which serve as targets (see Fig. 11, left).

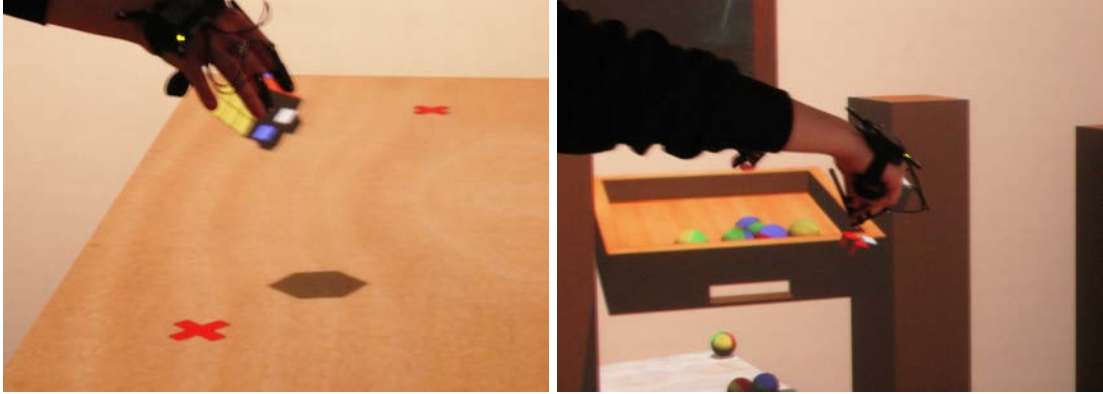


Figure 11: Left: In the training session, target crosses are used to guide the participants. Right: Usability task scene contains posts: the user passes the tray through them.

The experimenter demonstrates the movements in the IC before the training session starts, explains that an alarm sounds when the tray or hands hit the posts, and asks the user to test this. The goal of keeping as many balls as possible on the tray is explained, and that the participant should avoid hitting the posts with the tray or hands. The participant starts the training by lifting the tray off the table. The tray must then be rotated by 90 degrees, passed between the two first posts, rotated again by 90 degrees and passed between the two second posts before being released on one of the tables (preferably the one on the right). The user can then repeat these steps and try other movements until she feels comfortable with manipulating the tray. Time and trials are not restricted during the training session.

Once this is done, the user ends the session by placing the colored cube onto the red crosses in a specific order. The experimenter explains that the sound heard when the cube touches the marks means that the subtask is validated, and that the same sound will be used when validating a subtask in the following session.

Once the session is complete, the scene is reset so that the user can train with the other technique (wand or DM), and when it is complete for the second time, the next session is automatically loaded.

The tray and the cube are manipulated with the wand and one hand; the tray is also manipulated with two hands.

Usability Task:

We evaluate usability in a single task which tests all the manipulations.

The main evaluation scenario takes place in the same virtual room as in the training session, only this room contains a stool on the left, the same table on the right, the same three cardboard posts in between, and a cupboard in the back. Again, the virtual scene has the exact size of our four-sided IC. We place a tray with nine balls on the stool. There is also a bowl on the front half of the table (see Fig. 12, left).

This session is also performed in real conditions: the same scene has been built in the vicinity of the IC (see Fig. 12, right).

The experimenter first explains the entire task to the user by mimicking the required operations at the beginning of each condition, insisting on the order of the steps. The beginning and end of each subtask is determined automatically by the system. The task is composed as follows (see Fig. 13):



Figure 12: Left: Virtual Usability Environment. Right: Real Usability Space.

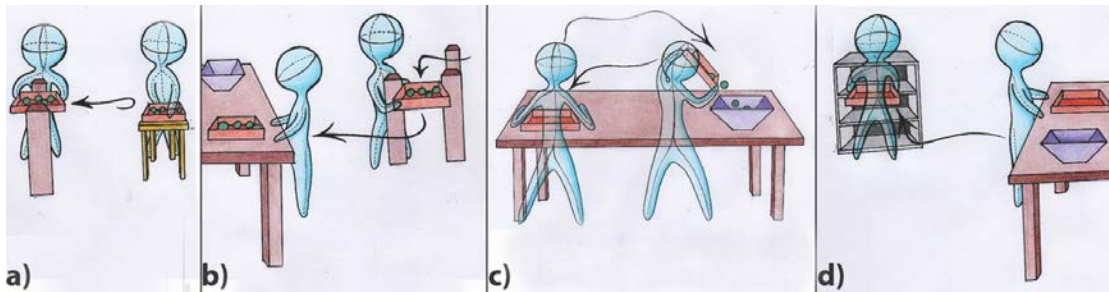


Figure 13: Subtasks of the Usability Task (see text).

- *Subtask 1:* The user grasps the tray and passes between the two sets of posts after rotating the tray by 90 degrees each time, and then releases it on the table. The participant is instructed to avoid dropping balls as well as to avoid touching the posts with the tray or the hands (see Fig. 13(a) and (b) and also Fig. 11, right).
- *Subtask 2:* The user lifts the tray off the table and empties it into the bowl before releasing it on the table again. The user is instructed to keep as many balls as possible into the bowl (Fig. 13(c)). If the tray is empty at the end of subtask 1, this subtask is skipped.
- *Subtask 3:* The user picks up the empty tray again and places it inside the cupboard. (Fig. 13(d)).

This task is quite challenging, even in the real world. Notice that the first subtask corresponds to what the user already had to do in the training session, as this is the most challenging part of the task.

Making Real and Virtual Equally Difficult:

To compare real and virtual tasks, we need to have approximately the same level of difficulty between the two. However, the physics simulator is only approximate in terms of material properties (friction etc.) and handles dynamics with impulses which can sometimes be unrealistic.

To perform a fair comparison, we performed a pilot study with four participants who performed the task of balancing the tray avoiding the posts and minimizing the number of balls dropped, in both the real and virtual environment. We measured difficulty by counting the number of balls dropped. The parameters we adjusted are the height of the borders of the virtual tray and the type of the real balls.

Because of the impulses, virtual balls tend to bounce more on the tray in the virtual environment than those in the real world. We thus increased the borders of the virtual tray. In addition, the real balls we used had different friction so we mixed two ball types: ping-pong balls and foam balls. The pilot study showed that equivalent difficulty level between real and virtual tasks is obtained with a tray border raised to $3/4$ the height of the balls, and mixing 3 foam balls and 6 ping-pong balls. We used this configuration in all experiments.

User Experience:

The free-form user experience session corresponds to a qualitative observational evaluation, to see what the user will do with almost no instruction.

The scene is a dining room, with a cupboard, shelves and a table. The table is empty, and plates, glasses, forks and knives for six people are placed in the cupboard on the left and on the shelves in the back (see Fig. 14).



Figure 14: : The user experience scene: Left, at the outset the table is clear. Right, at the end the table is set with plates, cups and cutlery.

The user is introduced to the scene and receives no other instructions than “you have 4 minutes to set the table as best as you can.” Our goal is to observe general behavior, i.e., whether participants use one or two hands, whether they focus on completely setting the table or correctly placing objects, the order of setting the table etc. We are interested in observing whether participants behave naturally, e.g., walk around the virtual table, catch falling objects etc., as well as their degree of presence.

4.4 Measurements

In each session, we record head position and orientation of the users at each frame. We also record the position and orientation of the fingers and palm, as well as the time to complete each task and subtask. Every object collision and ball dropped are also recorded. We videotaped the sessions, and manually extracted the completion times for the real environment.

Objective metrics:

During the usability task, we measure accuracy by recording the following: (1) position of the tray with respect to the posts to make sure that the user does pass the tray between them; (2) number of times the tray touches the posts; (3) number of times the hands touch the posts; (4) number of balls remaining on the tray when releasing the tray onto the table; (5) number of balls inside the bowl after emptying the tray.

Subjective measurements:

At the end of the experiment, participants complete a questionnaire. For each technique in virtual conditions (wand, one hand, and two hands), they are asked to rate various criteria on a Likert scale between 1 and 7. We evaluate: ease of use, fatigue caused by using the technique, sensation of “being there”, plausibility of the interaction with the environment and reaction of the environment to actions, similarity to the real condition (for one hand and two hands), precision, naturalness and cybersickness. Participants are also asked to rate the similarity of each virtual task to the real task.

4.5 Results

We next present the statistical analysis of the results for both our objective measurements, i.e., speed and errors and the responses to our subjective questionnaire. For the completion times, we performed a Shapiro test that rejected the normality hypothesis on the data distribution. Thus, we used a non-parametric Friedman test for differences among the conditions. Post-hoc comparisons were performed using Wilcoxon signed-rank tests with a threshold of 0.05 for significance.

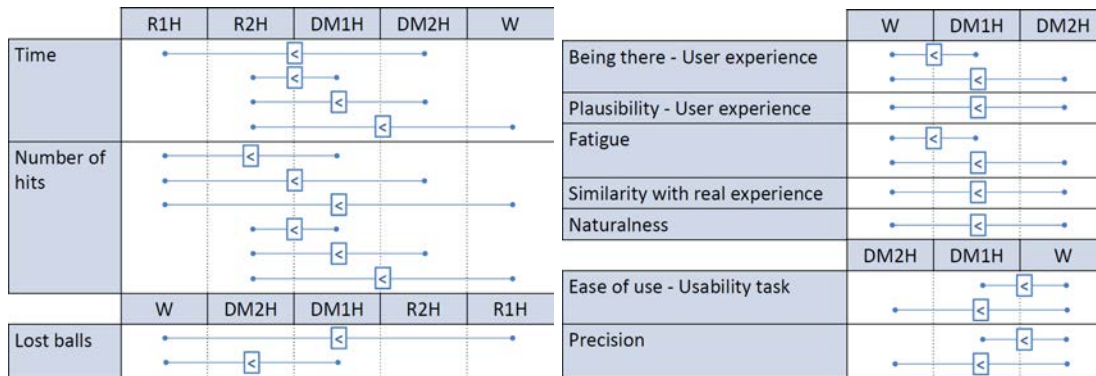


Figure 15: Left: significant “lower than” relationships between conditions for objective measurements. Right: Significant “lower than” relationships between conditions for subjective measurements.

Objective Measurements: We first performed a Friedman test on the time performances between the 5 conditions, i.e., 1- and 2-handed real, 1- and 2-handed DM and the wand. We use the following abbreviations: R1H and R2H are real, one and two handed conditions respectively, and DM1H and DM2H are virtual DM, one and two handed conditions respectively. The reported p-values were adjusted for multiple comparisons. We found a significant effect ($\chi^2 = 4.62$, $p < 0.001$) of condition. Post-hoc analysis revealed that the time to complete the task was significantly lower for R1H with a median time to completion of 61.33s compared to DM2H where

time to completion was 113.8s, ($p < 0.001$). The time for R2H (median=65.67s) was significantly lower than DM1H (median=86.95s, $p = 0.04$), DM2H ($p < 0.001$) and Wand (median=88.86s, $p = 0.04$). No significant effect was found between virtual conditions.

Errors were measured in two manners: first the number of balls lost throughout the task, and second the number of times the tray or the hands hit the posts in the virtual tasks. We also compare real and virtual conditions for both criteria.

We performed a Friedman test for the number of lost balls during the task. We found a significant effect of condition ($\chi^2 = 3.35$, $p = 0.007$). The number of lost balls was significantly lower for Wand compared to R1H and DM1H ($p = 0.007$ and $p = 0.01$ respectively), as revealed by post-hoc analysis.

Finally, we performed a Friedman test for the number of hits during the task. The test revealed a significant effect of condition ($\chi^2 = 5.32$, $p < 0.001$). The number of hits was significantly lower for R1H compared to all virtual conditions as shown by post-hoc analysis ($p < 0.001$ for DM1H, $p = 0.02$ for DM2H and $p < 0.001$ for Wand). The number of hits was also significantly lower for R2H compared to all virtual conditions ($p < 0.001$ for DM1H, $p = 0.003$ for DM2H and $p < 0.001$ for Wand). We did not find any significant effect between the virtual conditions.

Those results are summarized in Fig. 15(left) which shows “lower than” relationships between conditions for each parameter when it is significant.

Subjective Questionnaire: For the responses to the subjective questionnaire, we performed a Friedman test for the different criteria between the three virtual conditions. We did not find any significant effect for *Plausibility* for the usability task, *Being there* for the user experience and *Cybersickness*. We found a significant effect for 7 criteria: *Ease of Use* for the usability task ($\chi^2 = 4.95$, $p < 0.001$), *Being there* for the user experience ($\chi^2 = 2.69$, $p = 0.02$), *Plausibility* for the user experience ($\chi^2 = 2.54$, $p = 0.03$), *Fatigue* ($\chi^2 = 4.27$, $p < 0.001$), *Similarity with real experience* ($\chi^2 = 3.24$, $p = 0.003$), *Precision* ($\chi^2 = 4.35$, $p < 0.001$) and *Naturalness* ($\chi^2 = 2.70$, $p = 0.02$) (See Fig. 16). Post-hoc analysis showed that Wand was preferred to DM1H and DM2H for *Ease of Use* during the usability task ($p < 0.001$ and $p = 0.003$ respectively), *Fatigue* ($p < 0.001$ and $p = 0.007$ respectively) and *Precision* ($p < 0.001$ for both). In contrast, DM1H and DM2H conditions were significantly better rated than Wand for *Being there* for the user experience ($p = 0.04$ and $p = 0.02$ respectively). The Two Hands condition was preferred to Wand for *Plausibility* for the user experience ($p = 0.03$), *Similarity with real experience* ($p = 0.003$) and *Naturalness* ($p = 0.02$).

Those results are summarized in Fig. 15(right) which shows “lower than” relationships between conditions for each parameter when it is significant.

5 Discussion

Our goals were to evaluate the feasibility of DM in a fully immersive space, and its effect on presence as well as the similarity to real-world manipulation. We hypothesized that the wand would be more precise and efficient than DM, but we believed that DM would positively affect the sense of presence. The experimental results support these hypotheses.

Concerning the first hypothesis, the results show that all virtual tasks took longer than the real world task and that the wand and DM are equivalent in terms of speed (even though the Two Handed DM condition had a longer median completion time). We take this as an encouraging indication that DM does not penalize speed. However, as a general remark, it also indicates that for such complex tasks involving balance and rotations, we are still not at the point where virtual

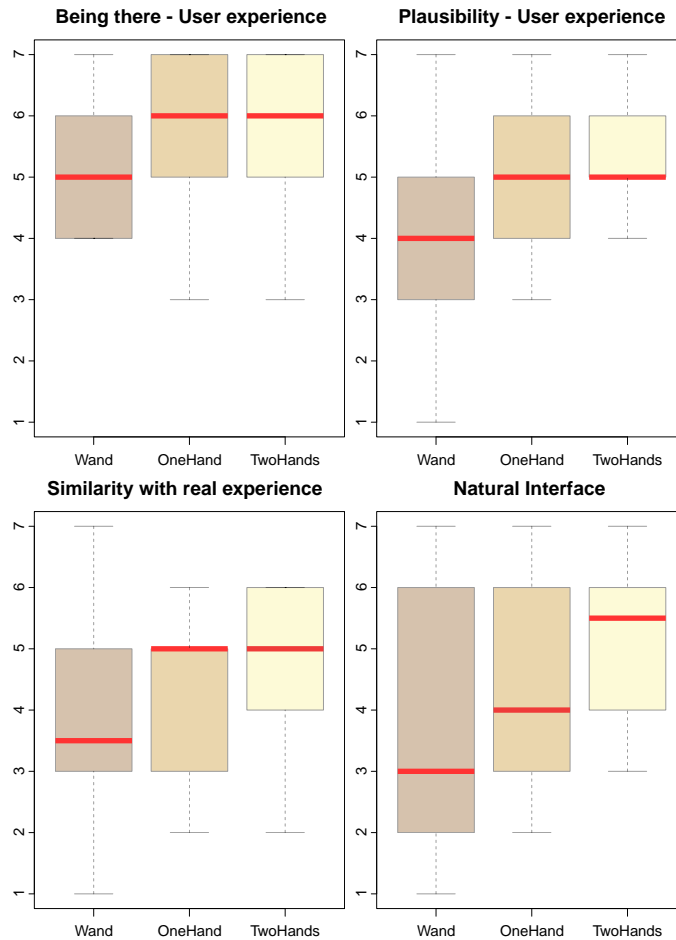


Figure 16: Boxplots of the statistical results for some of the most interesting studied criteria. Each boxplot is delimited by the 25% quantile and 75% quantile of the distribution of the effect over the individuals. The median is also represented as a red line for each effect.

tasks can be performed at the same speed as their real equivalent. In the real scene, the user is influenced by the haptic sense of weight which guides balance of the tray but increases fatigue. In contrast, this sense is missing in the virtual setting; note however that current solutions for haptics in immersive cubes are unsatisfactory and we thus chose not to opt for such a solution.

In terms of accuracy, users dropped fewer balls using the wand than with the one-handed virtual DM, as well as with the one-handed *real* life condition. There was no significant difference with the two-handed DM, real or virtual. Evidently, the wand is a “hyper-natural” interface in the terminology of Bowman et al. [5], so it is unsurprising that it allows better performance than the real world in some cases. The fact that this occurred for the one-handed case rather than two-hands is due to the inherent difficulty of the one-handed condition: the tray is a relatively long object, and a small movement of the fingers of the holding hand can result in a large movement of the tray, and thus a loss of the balls. When using two hands, the tray is more stable. The above observations are true for the virtual setting, but also for the real tasks; several participants

complained that the one-handed *real* task was hard and tiring.

The above two results for speed and accuracy indicate that the virtual DM is a feasible alternative for interaction. There is also an indication that DM could be considered better for applications such as training since performance is closer to the real world than the wand which *augments* interaction capabilities of the user.

The participants found the wand easier to use and less tiring. We also noticed that users subjectively considered the wand to be more precise than the DM even if the objective performance did not always confirm this, notably in terms of hitting the posts. Again, the “hyper-natural” aspect of the wand is a plausible explanation for this perception and this discrepancy.

For the user experience, participants rated the sense of *being there* to be higher for hands (both one and two-handed) compared to the wand. In informal interviews after completing the study, participants explained that in the usability task they were so concentrated on completing the task that they did not pay much attention to the environment; this was not the case however for the user experience, where there were no constraints. Similarly, two hands were rated higher than the wand for the sense of *plausibility*, again in the user experience. We believe that this is not the case for the one-hand case because of the lack of precision when manipulating objects, as explained above.

The above results on the subjective ratings show that our DM interface does have an effect on the two components of presence, *plausibility* and *being there* [23]. In addition, for the two-handed case, participants perceived them as more natural and closer to the real experience than the wand. We believe that these are encouraging results, which support our hypothesis that DM can improve the sense of immersion in such virtual environments and provide an experience that is closer to reality than using more traditional device-based interfaces.

We also observed participants behavior informally, which revealed several interesting cases of their reactions to our VE. Some of these are illustrated in the sequences of the accompanying video. The IC offers a high level of immersion in and of itself, for example participants attempted to place the wand on the virtual tables at the end of the different sessions. However, users tended to avoid walking through the virtual objects when they used their hands. In several cases, participants tried to catch the dropping objects as an automatic reaction when using DM. During the user experience, participants tended to place the objects in a specific order, as they do in real life: they begin with the plates, then the glasses, the forks, and they finish with the knives. Participants also tend to develop strategies to use the different techniques in the user experience. When using the wand, participants would stay in a convenient place to pick and place objects from a distance. However, they found it harder to rotate the objects, although different techniques for rotation could be implemented to alleviate this problem. When using DM, they used both hands to adjust the orientation of the plates; otherwise, they used both hands to pick two objects separately, and thus can set the table faster. These behaviors witness the immersion of our VE, as people behave as they do in real life.

Our results show that, even if the wand outperforms DM in terms of usability, our finger-based DM interface conveys a high sense of presence and naturalness. We also believe that improved tracking technology and possibly the use of 5 fingers will allow our approach to outperform the wand, while increasing difference in presence between the two.

6 Conclusions and Future Work

In this paper, we have presented a complete system which integrates direct manipulation with real-time physics in a fully immersive space, allowing close-to-natural manipulation of objects in a VE. Interaction is based on the physics engine, enhanced by a heuristic approach to manipulate

objects. Users can thus perform moderately complex tasks, involving translations and rotations of objects and maintaining balance while walking.

We performed an exploratory user study, which included a usability task, and a free-form user experience task. We tested our DM compared to the traditional wand, and for the controlled setting, we replicated the virtual task in the real world. Both the objective measures (speed, accuracy) and the responses to the subjective questionnaire indicated that DM is a feasible alternative to the more traditional wand interface. The results of our study also indicate that in several cases, especially when using two hands, the use of DM enhances the sense of presence in the virtual environment, and is perceived as being closer to reality.

In this study we examined feasibility of DM, leading us to test relatively complex tasks. Given that the feasibility of the interaction with DM is now quite clear, it will be interesting to examine the different parameters of our system in separate, more specific studies. An interesting direction to future work would also involve the use of a full hand model with complete tracking, similar to [10], which however requires improvements in depth sensor technology before becoming realisable.

7 Acknowledgements

The authors wish to thank Rachid Guerchouche for his help in many aspects of the project, also Theophanis Tsandilas, Martin Hachet, Anatole Lecuyer, Peter Vangorp and Sylvain Duchêne for their help and their useful comments. We also thank all the participants of the study. This work was supported in part by the Regional Council of Provence Alpes-Côte d’Azur and the EU project VERVE (<http://www.verveconsortium.eu>). The textures in the virtual environments come from the following websites: <http://archivetextures.net>, <http://www.freestockphotos.biz>, <http://www.flickr.com>, <http://www.defcon-x.de>, <http://www.public-domain-image.com>, <http://images.meredith.com>, <http://onlyhdwallpapers.com>, <http://www.merveilles-russie.com> and <http://www.zastavki.com>.

References

- [1] Anand Agarawala and Ravin Balakrishnan. Keepin’ it real: pushing the desktop metaphor with physics, piles and the pen. *CHI ’06*, pages 1283–1292, 2006.
- [2] Jacopo Aleotti and Stefano Caselli. Physics-based virtual reality for task learning and intelligent disassembly planning. *Virtual Reality*, 15:41–54, 2011.
- [3] C.W. Borst and A.P. Indugula. Realistic virtual grasping. In *IEEE VR 2005*, pages 91–98, 2005.
- [4] D.A. Bowman, S. Coquillart, B. Froehlich, M. Hirose, Y. Kitamura, K. Kiyokawa, and W. Stuerzlinger. 3d user interfaces: New directions and perspectives. *Computer Graphics and Applications, IEEE*, 28(6):20–36, 2008.
- [5] D.A. Bowman, R.P. McMahan, and E.D. Ragan. Questioning naturalism in 3d user interfaces. *Communications of the ACM*, 55(9):78–88, 2012.
- [6] Volkert Buchmann, Stephen Violich, Mark Billinghurst, and Andy Cockburn. Fingertips: gesture based direct manipulation in augmented reality. In *Proc. GRAPHITE ’04*, pages 212–221, 2004.

- [7] Marcio C. Cabral, Carlos H. Morimoto, and Marcelo K. Zuffo. On the usability of gesture interfaces in virtual reality environments. In *Proc. CLIHC '05*, pages 100–108, 2005.
- [8] Bernd Fröhlich, Henrik Tramberend, Andrew Beers, Maneesh Agrawala, and David Baraff. Physically-Based Manipulation on the Responsive Workbench. In *IEEE VR 2000*, 2000.
- [9] G. Heumer, H.B. Amor, M. Weber, and B. Jung. Grasp recognition with uncalibrated data gloves - a comparison of classification methods. In *Virtual Reality Conference, 2007. VR '07. IEEE*, pages 19–26, march 2007.
- [10] Otmar Hilliges, David Kim, Shahram Izadi, Malte Weiss, and Andrew Wilson. HoloDesk : Direct 3D Interactions with a Situated See-Through Display. In *CHI '12*, pages 2421–2430, 2012.
- [11] K. Hirota and M. Hirose. Dexterous object manipulation based on collision response. In *IEEE VR '03*, volume 2003, pages 232–239. IEEE Comput. Soc, 2003.
- [12] Daniel Holz, Sebastian Ullrich, Marc Wolter, and Torsten Kuhlen. Multi-Contact Grasp Interaction for Virtual Environments. *Journal of Virtual Reality and Broadcasting*, 5(7), 2008.
- [13] J. Jacobs and B. Froehlich. A soft hand model for physically-based manipulation of virtual objects. In *IEEE VR 2011*. IEEE, 2011.
- [14] Jan Jacobs, Michael Stengel, and Bernd Froehlich. A generalized God-object method for plausible finger-based interactions in virtual environments. In *3DUI'2012*, pages 43–51. Ieee, March 2012.
- [15] David B. Koons and Carlton J. Sparrell. Iconic: speech and depictive gestures at the human-machine interface. In *Conference Companion on Human Factors in Computing Systems, CHI '94*, pages 453–454, New York, NY, USA, 1994. ACM.
- [16] Marc Erich Latoschik. A gesture processing framework for multimodal interaction in virtual reality. In *Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation, AFRIGRAPH '01*, pages 95–100, New York, NY, USA, 2001. ACM.
- [17] M.E. Latoschik, M. Frohlich, B. Jung, and I. Wachsmuth. Utilize speech and gestures to realize natural interaction in a virtual environment. In *Industrial Electronics Society, 1998. IECON '98. Proceedings of the 24th Annual Conference of the IEEE*, volume 4, pages 2028–2033 vol.4, aug-4 sep 1998.
- [18] R.P. McMahan, A.J.D. Alon, S. Lazem, R.J. Beaton, D. Machaj, M. Schaefer, M.G. Silva, A. Leal, R. Hagan, and D.A. Bowman. Evaluating natural interaction techniques in video games. In *3DUI*, pages 11–14. IEEE, 2010.
- [19] M. Moehring and B. Froehlich. Natural interaction metaphors for functional validations of virtual car models. *IEEE TVCG*, 17(9), 2011.
- [20] R.G. O'Hagan, A. Zelinsky, and S. Rougeaux. Visual gesture interfaces for virtual environments. *Interacting with Computers*, 14(3):231 – 250, 2002.
- [21] Michael Ortega, Stephane Redon, and Sabine Coquillart. A six degree-of-freedom god-object method for haptic display of rigid bodies with surface properties. *IEEE TVCG*, 13(3):458–469, May 2007.

- [22] M. Prachyabrued and C.W. Borst. Visual interpenetration tradeoffs in whole-hand virtual grasping. In *3DUI*, pages 39–42. IEEE, 2012.
- [23] M. Slater. Place illusion and plausibility can lead to realistic behaviour in immersive virtual environments. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535):3549–3557, 2009.
- [24] D. J. Sturman, D. Zeltzer, and S. Pieper. Hands-on interaction with virtual environments. *UIST '89*, pages 19–24, 1989.
- [25] Thomas Ullmann and Joerg Sauer. Intuitive Virtual Grasping for non Haptic Environments. In *Pacific Graphics '00*, pages 373–381, 2000.
- [26] Alan Wexelblat. An approach to natural gesture in virtual environments. *ACM Trans. Comput.-Hum. Interact.*, 2(3):179–200, September 1995.
- [27] Andrew D. Wilson, Shahram Izadi, Otmar Hilliges, Armando Garcia-Mendoza, and David Kirk. Bringing Physics to the Surface. In *ACM UIST '08*, pages 67–76, 2008.



**RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE**

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-0803