Virtual Reality manuscript No.
(will be inserted by the editor)

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

Received: date / Accepted: date

Abstract A solution for interaction using finger tracking in a cubic immersive virtual reality system (or immersive cube) is presented. 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. A first study has been performed to evaluate our interface, with tasks involving complex manipulations, such as balancing objects while walking in the cube. The users finger-tracked manipulation was compared to manipulation with a 6 degree-of-freedom 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.

E. Chapoulie · G. Drettakis
Inria, REVES, Sophia Antipolis, France
E-mail: george.drettakis@inria.fr

M. Marchal
Inria, Hybrid, Rennes, France

M. Marchal
IRISA-INSA, Rennes, France

E. Dimara · M. Roussou
University of Athens, Athens, Greece

M. Roussou
Makebelieve Design and Consulting, Athens, Greece

J.-C. Lombardo
Inria, Sophia Antipolis, France

Keywords Virtual reality · Direct manipulation · Immersive cube · Finger tracking

1 Introduction

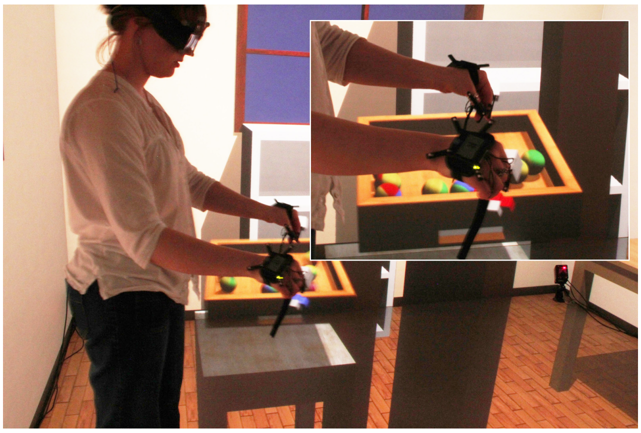

Fig. 1 A user in the four-sided cube holding a tray with two hands.

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 since they allow users to perform actions which cannot be performed naturally, such as picking objects from afar, or navigating while physically staying in the same place. However, one goal of fully immersive systems is to enhance presence and immersion [Slater (2009)]. In such a context, flysticks can potentially degrade the realism and naturalness of the virtual environment (VE). To avoid this shortcoming, the use of direct manipulation

(DM) using finger tracking as close as possible to the manipulation used in the real world is investigated [Moehring and Froehlich (2011); Wexelblat (1995); Jacobs et al (2012); Hilliges et al (2012)].

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 [Jacobs et al (2012); Moehring and Froehlich (2011)], 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.

A solution which incorporates DM for grasping and moving objects, using both one and two hands, based on finger tracking with a glove-like device is presented and evaluated. 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. A physics engine, or simulator, is a software library that approximates the physical behavior of virtual objects by computing collision detection, the effect of gravity, etc., and the corresponding object transformations. It allows close-to-natural manipulation of objects in the scene. An approach for DM which is robust to tracker noise and instabilities of the physics simulation is presented.

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. DM is compared to traditional wand-based interaction and, most importantly, compared to a real-world reproduction of the virtual task, through a study focused on the sense of immersion. We record both objective measurements (time to completion and errors) and subjective judgments through the use of a questionnaire.

In summary, 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, clearly demonstrating its utility. 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.

2 Previous work

Gesture-based interaction has received significant interest in virtual or augmented reality research [Sturman et al (1989); O'Hagan et al (2002); Buchmann et al (2004); Cabral et al (2005)]. A thorough review of natural gestures for virtual reality (VR) [Bowman et al (2012)] 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) [Bowman et al (2008)], 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 [Buchmann et al (2004)] even if the number of gestures is often limited [Sturman et al (1989); O'Hagan et al (2002); Cabral et al (2005)]. The early work by Bolt (1980) already investigates the combined use of gestures and voice inputs through a set of commands to manipulate simple shapes. However, the use of a specific vocabulary 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 [Koons and Sparrell (1994); Latoschik et al (1998); Latoschik (2001)]. We are more interested in the case of DM, i.e., interacting with the environment in a natural manner with the users hands [Moehring and Froehlich (2011); Wexelblat (1995); Jacobs et al (2012); Hilliges et al (2012)].

Using the hand as an interaction metaphor results from an interest in applying the skills, dexterity and naturalness of the human hand directly to humancomputer interfaces. Such metaphors are typically achieved either by detecting the hand through computer vision-based algorithms or by wearing-specific devices such as gloves [Sturman and Zeltzer (1994)]. However, Wang and Popović (2009) recently proposed a solution combining both technologies to provide simple and accurate real-time hand tracking for desktop VR applications, using only one camera and a color glove with a specific pattern. They demonstrate the validity of their solution through typical applications such as virtual assembly or gesture recognition. We are interested in fully immersive setups, and thus will focus on finger tracking technologies with high-range detection. Sturman and Zeltzer (1994) and Dipietro et al (2008) provide thorough surveys of such devices and their applications in various fields, such as design for construction or 3D modeling, data visualization, robot control, entertainment and sign language interpretation, while the health sector shows an increasing interest in such interfaces for motor rehabilitation (hand functional assessment), ergonomics or training.

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 [Agarawala and Balakrishnan (2006); Wilson et al (2008)]. 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 Fröhlich et al (2000), 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, sometimes referred to as kinematic objects; they can apply forces to dynamic objects in the scene, but no object can affect them) with respect to the simulation of the rest of the environment, i.e., correctly providing external forces from the hands.

Over the last 10 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 [Ortega et al (2007)]. Initial work, before physics simulation became widely available, focused on the definition of appropriate gestures for certain kinds of operations, and notably grasping [Ullmann and Sauer (2000)]. Experiments with force-feedback systems allowed the simulation of several quite complex manipulations [Hirota and Hirose (2003)]. A spring model coupled with a commercially available physics simulator was used in Borst and Indugula (2005) to simulate various kinds of grasping operations with objects of varying complexity and to avoid hand-object interpenetrations. A simpler approach was proposed by Holz et al (2008), where grasping is simulated without complex physics. Grasping and interpenetration were also the focus of the work by Prachyabrued and Borst (2012).

In Moehring and Froehlich (2011), DM was used based on grasping heuristics and constraint-based solutions for the specific case of a car interior. Their approach is based on an analysis of the types of objects, their constraints and the typical grasps to derive a set of pseudophysical interaction metaphors. A quantitative comparison to a real-world car interior (mirror, door, etc.) was performed. More recent work has concentrated on developing appropriate soft models [Jacobs and Froehlich (2011)] and efficient solvers to avoid interpenetration of God-objects and other objects in the scene [Jacobs et al (2012)], 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 [Hilliges et al (2012)]. 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 [Aleotti and Caselli (2011)] or the use of natural interaction for video games [McMahan et al (2010)]. Heumer et al (2007) 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.

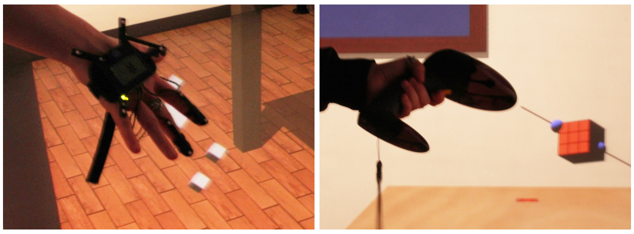

Fig. 2 *Left* visual representation of the palm and three fingertips in the VE (cubes are shown away from the fingers for clarity of illustration). *Right* wand selection used for comparison.

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 [Prachyabrued and Borst (2012); Hilliges et al (2012)], a room-sized environment is targeted, and the user is allowed to walk around the scene while manipulating objects. There are three main difficulties discussed below: finger tracking, dynamic constraints between objects and the physics simulation.

First, finger tracking in the IC is challenging. Even though a high-end finger tracking system is used (Sect. 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 1020 cm away from the screen. To overcome these issues, our tasks are designed to avoid activities where the users body blocks the tracking cameras or involves objects in the shadow regions. Our treatment of finger tracking and tracker signal filtering is detailed in Sect. 3.1.

Second, given our goal of assessing feasibility of DM, users are asked 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 developed.

Third, our tasks require the use of a fast physics engine; we use *Bullet*[1] (see Sect. 3.4 for details), which is fast enough to handle quite complex scenes in real time (e.g., our user experience scene in Fig. 14. Due to the low light condition in the IC, photos were taken with a long exposure which results in some blur in the photographs). *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 interpenetration of the hand and other objects in the scene [e.g., Jacobs et al (2012); Prachyabrued and Borst (2012)], while other approaches [Ullmann and Sauer (2000); Holz et al (2008)] avoid the need for complexand often expensiveprecise simulation using specific algorithms . In contrast, a fast but simple physics simulation is used, combined 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.). (Fig. 1).

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. The solutions adopted to address these problems are next discussed.

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). Cubes are used 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; these are called *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.

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”). A Kalman filter is applied to each finger in time and we 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

The physics engine is directly used 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, contact events with the fingers provided by the physics simulation are tracked. Once two kinematic contacts on a given object are identified, the object is marked as “grasped,” until the contacts are released. Grasped objects are removed from the physics simulation, and the transformation and speed of the trackers are enforced so that the simulation of objects in contact with the object selected is still correct.

When grasping, specific data are stored (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. These data are 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. The selection is made more robust by marking an object as “released” only when the distance between the contacts is $>10\%$ 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

[1] http://www.bulletphysics.org

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*. These thresholds have been chosen by pilot trial-and-error tests in several different settings. While releasing, the object being handled follows the translation of the midpoint 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.

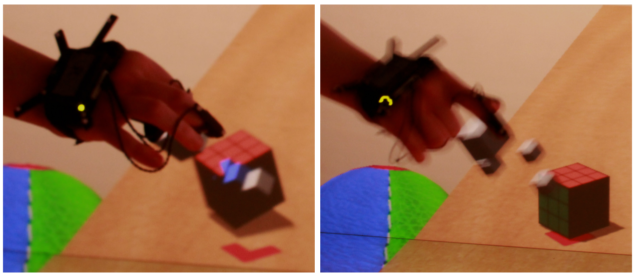

Fig. 3 Accurately placing a cube in a corner.

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 (Figs. 4, 5).

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 reinserted into the physics simulation.

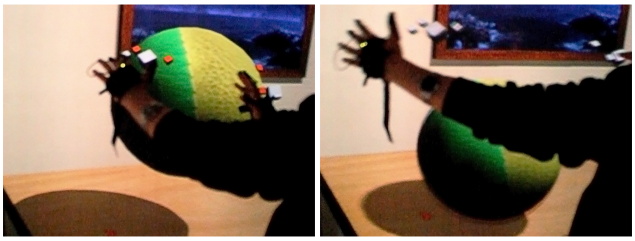

Fig. 4 User spreads her arms to release a large object without the need to further open her hands.

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, a finite-state machine approach is designed. The different possible transitions for each manipulation are next described.

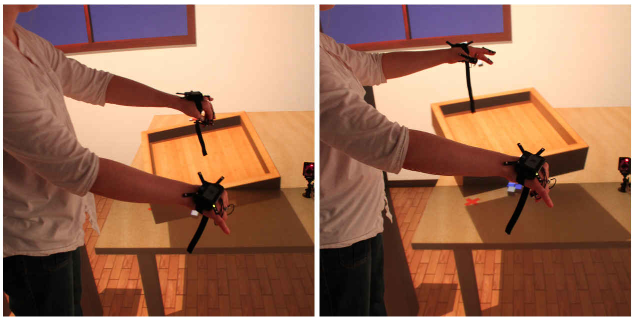

Fig. 6 *Left* two-handed grasp. *Right* two-handed release.

3.3.1 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.

3.3.2 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 two hands, the user must spread her 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.

3.3.3 Translate

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

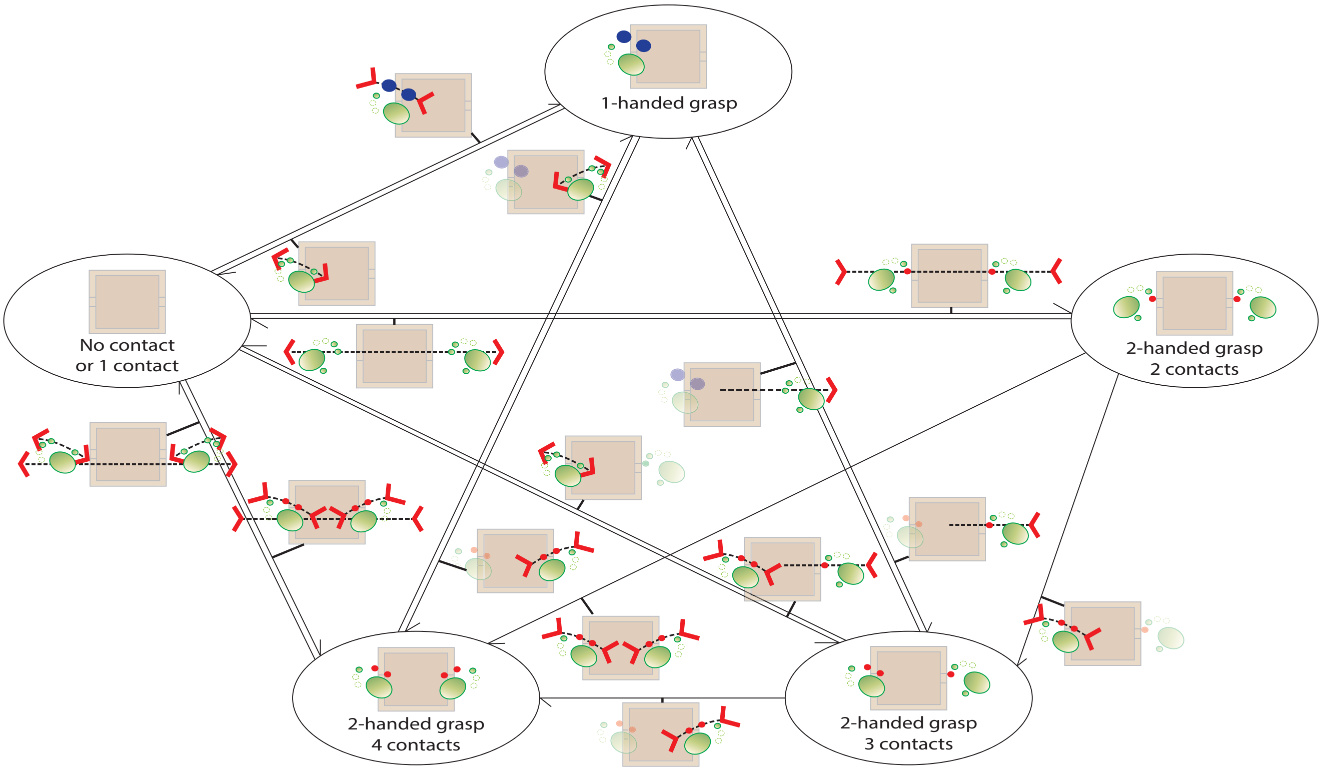

Fig. 5 Finite-state machine graph detailing transitions between grasping and releasing states. A representation of the hand is shown in *green*; *dots* represent fingers. When only one active finger is touching a dynamic object, the corresponding *dot* becomes *plain green*. When two active fingers of 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*.

3.3.4 Rotate

The grasped object is rotated as soon as the fingers in contact rotate. The rotation with two hands is currently limited 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).

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.

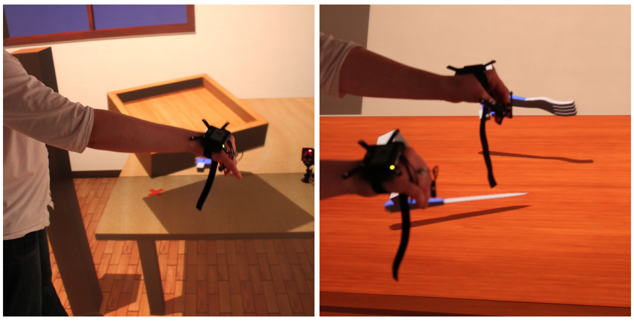

Fig. 7 *Left* one-handed grasp. *Right* two hands grasping two objects.

We use our in-house VR software system which is based on OpenSceneGraph.[2] The Bullet Physics open source physics engine has been adapted to communicate the data needed by our grasping/release approach and our finite-state machine through extra object properties, as the transformations of the handled objects are enforced coherently with those controlled by the physics simulation. The interocular distance is calibrated, as well as YZ 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

[2] http://www.openscenegraph.org/

the projection from the users point of view (Sect. 4.2). This is also more comfortable for the user as it is expected to reduce the risk of cybersickness.

4 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, a complex task has purposely been chosen, which involves quite complex translations and rotations, while walking and balancing objects on a tray (see Sect. 4.3 for details). By *complex task*, we mean a task requiring the users attention as it combines several aspects of everyday movements, such as accuracy and balance, such that the task is difficult to perform. We do this evaluation 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 as follows: 1) 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.

Both objective measurements, namely time to complete a task and precision or errors, and subjective judgments based on a questionnaire are recorded.

4.1 Population

The experiment has been run 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

A four-sided IC (the Barco iSpace [3] is used, comprising three walls and a floor, which has four retro-projected “black” screens. The front and side screens are 3.2 m wide 2.4 m high, and the floor is 3.2 m wide 2.4 m long. Stereo is provided using Infitec technology, and for tracking, an ART [4] infrared optical system is used, with eight cameras (see Fig. 8, left). The head is tracked 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.

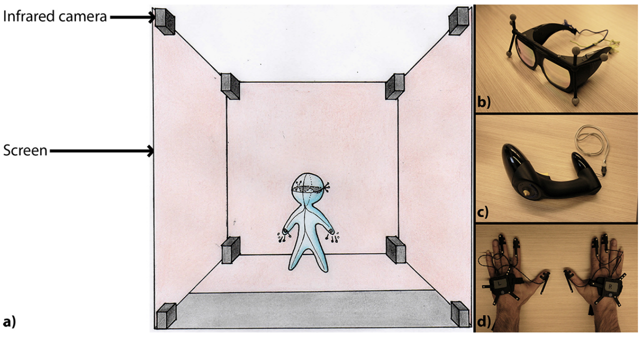

Fig. 8 Experimental setup. The four-sided IC is represented in a) with its eight infrared cameras. On the right are photos of the tracked input devices used: b) Infitec glasses with frame, c) ART flystick and d) ART finger-tracking gloves.

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 users immersive experience. Our scenarios were thus designed to avoid eye gaze on these areas as much as possible, and our models were also designed to hide the cables and the cameras by adding very dark baseboards to the room.

4.2.1 Wand interaction

For the wand, a standard virtual ray emanating from the wand is used. To grasp an object with the wand, the user points the wand toward 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.

The ray has been implemented to be thin to limit occlusion and to be long enough (1 m) to allow sufficient coverage of the 3.2 m 3.2 m 2.4 m IC.

4.3 Experimental procedures and environments

The experiment lasted 90 min on average. As every participant performed all the tasks, it is a within-subject design. The experiment consists of a calibration step, a training session, a usability test, a “free form” user experience and a

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

[4] http://www.ar-tracking.com

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.

4.3.1 Calibration step

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

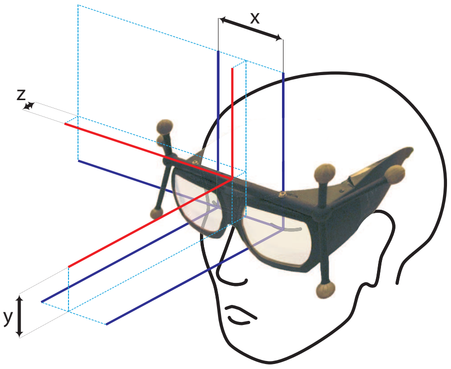

Fig. 9 Position of the user's eyes is calibrated in the three dimensions.

The position of the eyes is also calibrated. A pilot test has been performed where the interocular distance was simply measured and set: this proved to be insufficient since the environment still displayed a "swimming" effect. To overcome this, the other two coordinates of the eye positions are also adjusted (see Fig. 9). This is done using a simple scene, i.e., a floor and a stool (see 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.

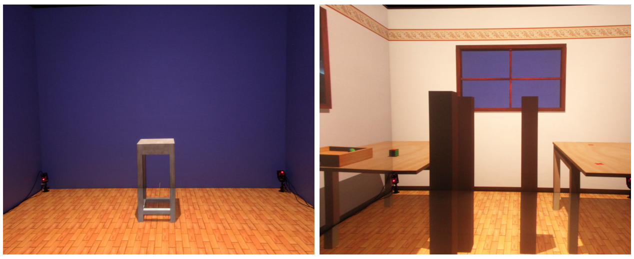

Fig. 10 *Left* calibration scene. *Right* training scene

4.3.2 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).

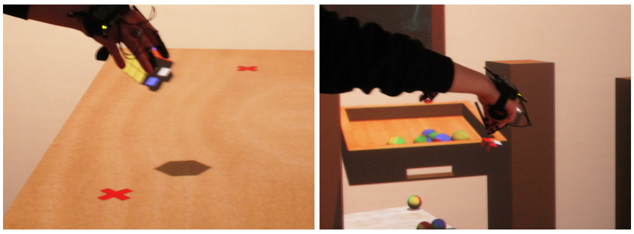

Fig. 11 *Left* in the training session, target crosses are used to guide the participants. *Right* usability task scene contains posts: the user passes the tray through them.

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°, passed between the two first posts, rotated again by 90° 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.

4.3.3 Usability task

Usability is evaluated 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. A tray with nine balls is placed 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).

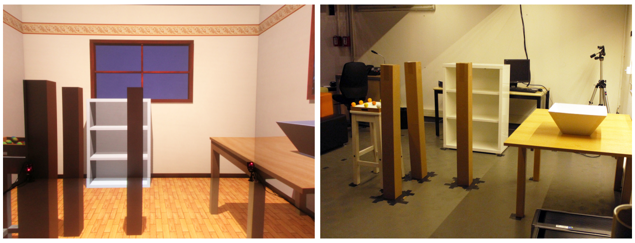

Fig. 12 *Left* virtual usability environment. *Right* real usability space.

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):

- *Subtask 1:* The user grasps the tray and passes between the two sets of posts after rotating the tray by 90° 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 Figs. 13 a, b, 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.

4.3.4 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. For a fair comparison, a pilot test with four participants who performed the task of balancing the tray avoiding the posts and minimizing the number of balls dropped has been carried out before the actual experiments, in both the real and virtual environments. Difficulty is measured by counting the number of balls dropped. The adjusted parameters 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 VE than those in the real world. The borders of the virtual tray have thus been increased. In addition, the real balls had different friction so two ball types were mixed: ping-pong balls and foam balls. The pilot test 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. This configuration has been used in all experiments.

4.3.5 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).

The user is introduced to the scene and receives no other instructions than “you have 4 min 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

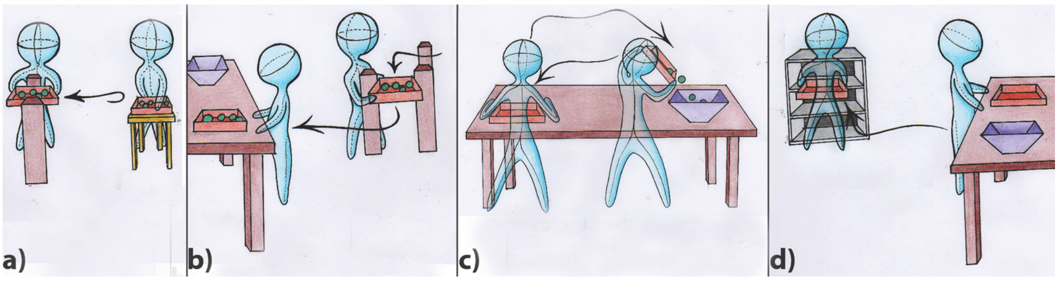

Fig. 13 Subtasks of the usability task (see text).

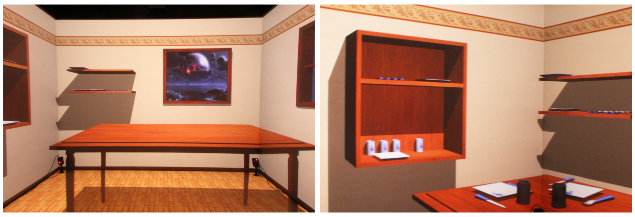

Fig. 14 User experience scene: *Left*; at the outset, the table is clear. *Right*; at the end, the table is set with plates, cups and cutlery.

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

4.4.1 Objective metrics

During the usability task, accuracy is measured 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.

4.4.2 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's 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. Then, they have to answer open questions related to their strategies for using the interfaces, and their opinion on advantages, drawbacks and difficulties of each interface. Finally, they are free to make additional comments. For further detail, please refer to the questionnaire in supplemental material.

4.5 Results

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

4.5.1 Objective measurements

A Friedman's test has first been performed on the time performances between the 5 conditions, i.e., 1- and 2-handed real, 1- and 2-handed DM and the wand. The following abbreviations are used: 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. A significant effect ($\chi^2 = 4.62$, $p < 0.001$) of condition has been found. Post hoc analysis revealed that the time to complete the task was significantly lower for R1H with a median time to completion of 61.33 s compared to DM2H where time to completion was 113.8 s ($p < 0.001$). The time for R2H (median = 65.67 s) was significantly lower than DM1H (median = 86.95 s, $p = 0.04$), DM2H ($p < 0.001$) and wand (median = 88.86 s, $p = 0.04$). No significant effect was found between virtual conditions.

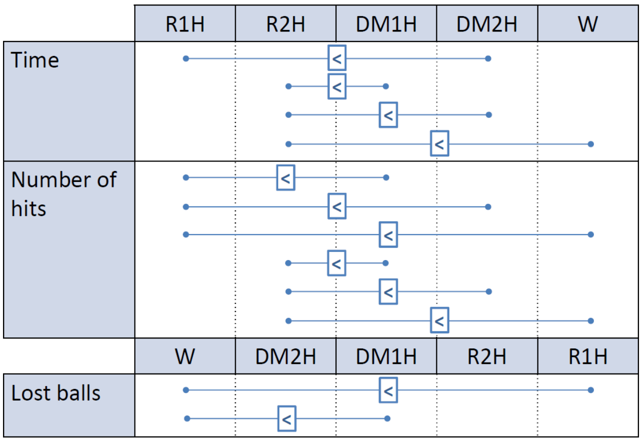

Fig. 15 Significant "lower than" relationships between conditions for objective measurements.

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. Real and virtual conditions for both criteria are also compared.

A Friedman's test has been performed for the number of lost balls during the task. A significant effect of condition ($\chi^2 = 3.35$, $p = 0.007$) has been found. 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, a Friedman's test has been performed 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). No significant effect has been found between the virtual conditions.

Those results are summarized in Fig. 15 which shows "lower than" relationships between conditions for each parameter when it is significant.

4.5.2 Subjective questionnaire

For the responses to the subjective questionnaire, a Friedman's test has been performed for the different criteria between the three virtual conditions. No significant effect has been found for *Plausibility* for the usability task, *Being there* for the user experience and *Cybersickness*. A significant effect has been found 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 Figs. 16, 17). 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* ($p = 0.02$).

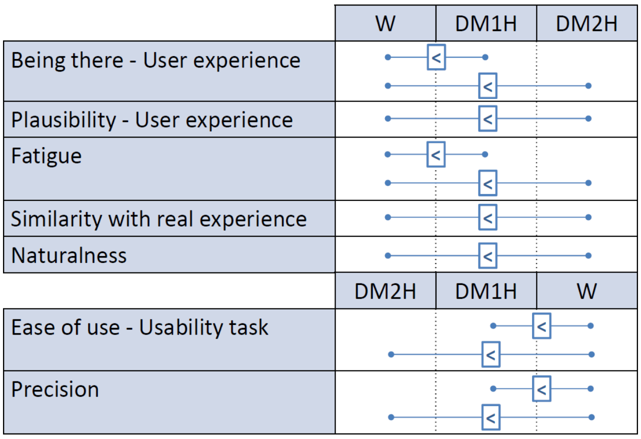

Fig. 16 Significant "lower than" relationships between conditions for subjective measurements.

Those results are summarized in Fig. 16 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

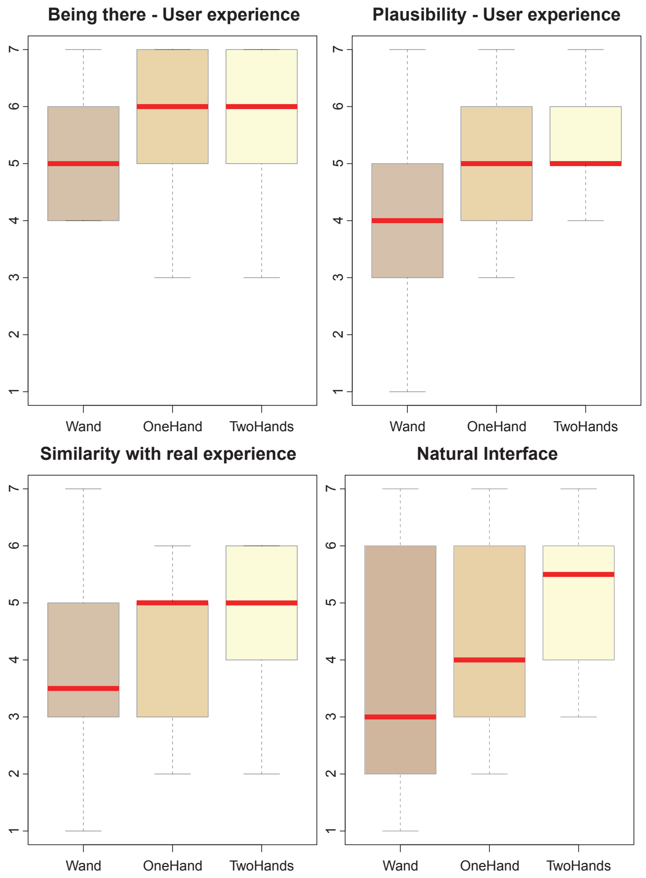

Fig. 17 *Boxplots* of the statistical results for some of the most interesting studied criteria. Each *boxplot* is delimited by the 25% quantile and 75% quantile of the distribution of the effect over the individuals. The median is also represented as a *red line* for each effect.

and rotations, we are still not at the point where virtual tasks can be performed at the same speed as their real equivalent. In the real scene, the user is influenced by the kinesthetic perception (of touch, weight and muscle tension) 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 ICs 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 (2012), 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* [Slater (2009)]. 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 VEs 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, and they finish with the forks and 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 five 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, a complete system has been presented which integrates DM 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.

A first user study has been performed, which included a usability task, and a free-form user experience task. Our DM has been compared to the traditional wand, and for the controlled setting, the virtual task has been replicated 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 VE and is perceived as being closer to reality.

In this study, feasibility of DM has been examined, 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 Hilliges et al (2012), which, however, requires improvements in depth sensor technology before becoming realizable.

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 dAzur and the EU project VERVE (http://www.verveconsortium.eu). The textures in the VEs 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

Agarawala A, Balakrishnan R (2006) Keepin' it real: pushing the desktop metaphor with physics, piles and the pen. CHI '06, pp 1283–1292

Aleotti J, Caselli S (2011) Physics-based virtual reality for task learning and intelligent disassembly planning. Virtual Reality 15:41–54

Bolt RA (1980) &ldquo;put-that-there&rdquo;: Voice and gesture at the graphics interface. In: Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, ACM, New York, NY, USA, SIGGRAPH '80, pp 262–270, DOI 10.1145/800250.807503, URL http://doi.acm.org/10.1145/800250.807503

Borst C, Indugula A (2005) Realistic virtual grasping. In: IEEE VR 2005, pp 91–98

Bowman D, Coquillart S, Froehlich B, Hirose M, Kitamura Y, Kiyokawa K, Stuerzlinger W (2008) 3d user interfaces: New directions and perspectives. Computer Graphics and Applications, IEEE 28(6):20–36

Bowman D, McMahan R, Ragan E (2012) Questioning naturalism in 3d user interfaces. Communications of the ACM 55(9):78–88

Buchmann V, Violich S, Billinghurst M, Cockburn A (2004) Fingartips: gesture based direct manipulation in augmented reality. In: Proc. GRAPHITE '04, pp 212–221

Cabral MC, Morimoto CH, Zuffo MK (2005) On the usability of gesture interfaces in virtual reality environments. In: Proc. CLIHC '05, pp 100–108

Dipietro L, Sabatini AM, Dario P (2008) A survey of glove-based systems and their applications. IEEE Transactions on Systems, Man, and Cybernetics, Part C 38(4):461–482

Fröhlich B, Tramberend H, Beers A, Agrawala M, Baraff D (2000) Physically-Based Manipulation on the Responsive Workbench. In: IEEE VR 2000

Heumer G, Amor H, Weber M, Jung B (2007) Grasp recognition with uncalibrated data gloves - a comparison of classification methods. In: Virtual Reality Conference, 2007. VR '07. IEEE, pp 19 –26, DOI 10.1109/VR.2007.352459

Hilliges O, Kim D, Izadi S, Weiss M, Wilson A (2012) HoloDesk : Direct 3D Interactions with a Situated See-Through Display. In: CHI '12, pp 2421–2430

Hirota K, Hirose M (2003) Dexterous object manipulation based on collision response. In: IEEE VR '03, IEEE Comput. Soc, vol 2003, pp 232–239

Holz D, Ullrich S, Wolter M, Kuhlen T (2008) Multi-Contact Grasp Interaction for Virtual Environments. Journal of Virtual Reality and Broadcasting 5(7):1860–2037

Jacobs J, Froehlich B (2011) A soft hand model for physically-based manipulation of virtual objects. In: IEEE VR 2011, IEEE

Jacobs J, Stengel M, Froehlich B (2012) A generalized God-object method for plausible finger-based interactions in virtual environments. In: 3DUI'2012, Ieee, pp 43–51

Koons DB, Sparrell CJ (1994) Iconic: speech and depictive gestures at the human-machine interface. In: Conference Companion on Human Factors in Computing Systems, ACM, New York, NY, USA, CHI '94, pp 453–454, DOI 10.1145/259963.260487, URL http://doi.acm.org/10.1145/259963.260487

Latoschik M, Frohlich M, Jung B, Wachsmuth I (1998) 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, vol 4, pp 2028 –2033 vol.4, DOI 10.1109/IECON.1998.724030

Latoschik ME (2001) A gesture processing framework for multimodal interaction in virtual reality. In: Proceedings of the 1st international conference on Computer graphics, virtual reality and visualisation, ACM, New York, NY, USA, AFRIGRAPH '01, pp 95–100, DOI 10.1145/513867.513888, URL http://doi.acm.org/10.1145/513867.513888

McMahan R, Alon A, Lazem S, Beaton R, Machaj D, Schaefer M, Silva M, Leal A, Hagan R, Bowman D (2010) Evaluating natural interaction techniques in video games. In: 3DUI, IEEE, pp 11–14

Moehring M, Froehlich B (2011) Natural interaction metaphors for functional validations of virtual car models. IEEE TVCG 17(9):1195–1208

O'Hagan R, Zelinsky A, Rougeaux S (2002) Visual gesture interfaces for virtual environments. Interacting with Computers 14(3):231 – 250

Ortega M, Redon S, Coquillart S (2007) A six degree-of-freedom god-object method for haptic display of rigid bodies with surface properties. IEEE TVCG 13(3):458–469

Prachyabrued M, Borst C (2012) Visual interpenetration tradeoffs in whole-hand virtual grasping. In: 3DUI, IEEE, pp 39–42

Slater M (2009) 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

Sturman DJ, Zeltzer D (1994) A survey of glove-based input. IEEE Comput Graph Appl 14(1):30–39, DOI 10.1109/38.250916, URL http://dx.doi.org/10.1109/38.250916

Sturman DJ, Zeltzer D, Pieper S (1989) Hands-on interaction with virtual environments. UIST '89, pp 19–24

Ullmann T, Sauer J (2000) Intuitive Virtual Grasping for non Haptic Environments. In: Pacific Graphics '00, pp 373–381

Wang RY, Popović J (2009) Real-time hand-tracking with a color glove. In: ACM SIGGRAPH 2009 Papers, ACM, New York, NY, USA, SIGGRAPH '09, pp 63:1–63:8, DOI 10.1145/1576246.1531369, URL http://doi.acm.org/10.1145/1576246.1531369

Wexelblat A (1995) An approach to natural gesture in virtual environments. ACM Trans Comput-Hum Interact 2(3):179–200

Wilson AD, Izadi S, Hilliges O, Garcia-Mendoza A, Kirk D (2008) Bringing Physics to the Surface. In: ACM UIST '08, pp 67–76