View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by SAM; the Arts et Métiers ParisTech open access repository

Science Arts & Métiers (SAM)

is an open access repository that collects the work of Arts et Métiers ParisTech researchers and makes it freely available over the web where possible.

This is an author-deposited version published in: https://sam.ensam.eu
Handle ID: .http://hdl.handle.net/10985/9055

To cite this version :

Franck HERNOUX, Olivier CHRISTMANN - A seamless solution for 3D real-time interaction: design and evaluation - Virtual Reality p.1-20 - 2014

Any correspondence concerning this service should be sent to the repository Administrator : archiveouverte@ensam.eu

Arts et Métiers Institute of Technology

DOI 10.1007/s10055-014-0255-z

A seamless solution for 3D real-time interaction: design and evaluation

Franck Hernoux · Olivier Christmann

Abstract This paper aims to propose and evaluate a markerless solution for capturing hand movements in real time to allow 3D interactions in virtual environments (VEs). Tools such as keyboard and mice are not enough for interacting in 3D VE; current motion capture systems are expensive and require wearing equipment. We developed a solution to allow more natural interactions with objects and VE for navigation and manipulation tasks. We conducted an experimental study involving 20 participants. The goal was to realize object manipulation (moving, orientation, scaling) and navigation tasks in VE. We compared our solution (Microsoft Kinect-based) with data gloves and magnetic sensors (3DGloves) regarding two criteria: performance and acceptability. Results demonstrate similar performance (precision, execution time) but a better overall acceptability for our solution. Preferences of participants are mostly in favor of the 3DCam, mainly for the criteria of comfort, freedom of movement, and handiness. Our solution can be considered as a real alternative to conventional systems for object manipulation in virtual reality.

Keywords Virtual reality · Seamless solution · Hand tracking · Real time · 3D interaction · 3D camera

The desktop metaphor, which appeared on Apple computers in 1984, was the real beginning of the Window Icon Menu Pointer paradigm (Beaudouin-Lafon 2004), which goes together with the use of the mouse. Since then, despite the fact that much effort has been focused on improving graphical user interfaces, interactions are still based on the keyboard–mouse duo. Nowadays, applications, either games or professional (e.g., modeling …), allow the user to navigate or to "manipulate" content in a tridimensional environment. In this case, interactions are done with six degrees of freedom (DoF) (three for position and three for orientation). Usually, these DoFs are factored into 2D subspaces that are mapped on the axis of a mouse (Wang et al. 2011), where metaphors help to facilitate understanding and interaction (e.g., manipulating an imaginary sphere for rotating an object (Chen et al. 1988). 2D devices are widespread but their performance or usability is debatable (Berard et al. 2009). In addition, current devices are sometimes unsuitable, for example in medicine where seamless devices are sometimes necessary (Fuchs and Mathieu 2003). In fact, no interactive device for manipulating 3D objects has been widely adopted by the general public. If advances regarding tangible user interactions seem promising (Poor et al. 2013), they impose to look away from the screen to watch the manipulated object.

Currently, most devices follow Norman's model (Norman 1988) (Fig. 1) where a user's stimuli are sent through any device to the computer that returns visual data to the user (black arrows). This model was enriched (gray dotted arrows) by (Nedel et al. 2003) with a feedback loop from the computer to the user through the device. A behavioral interface (Fuchs and Moreau 2003), which makes the interface "disappear" to allow a "natural" interaction for users, is often presented as the final outcome of direct manipulation interfaces; it could be the next step to improve Norman's model. The feeling of immersion and presence can be enhanced if the user has no equipment to wear; the system becomes seamless to the user (Winkler et al. 2007). Allowing the user to interact directly with his

F. Hernoux (✉) · O. Christmann
LAMPA - Arts & Métiers ParisTech,
2, Bd du Ronceray, 49000 Angers, France
e-mail: hernouxfranck@yahoo.fr

O. Christmann
e-mail: olivier.christmann@ensam.eu

or her hands in a virtual environment would be the best way to make the system seamless when manipulating content.

The hand provides 70 % of our motor skills (Lempereur 2008) and offers many different types of opposition-based grasps (Tubiana and Kapandji 1991), which allow us to manipulate and interact with our environment. It seems simple to use this natural interaction for implementing 3D interfaces but in fact the challenge is to transpose the manual interaction (or bimanual) from the real world to the virtual one, with complete transparency to the user. A great deal of research work has been done using stereoscopic systems [e.g., in computer-aided design (CAD) area (Wang et al. 2011)], but the emergence of 3D cameras (especially Microsoft's Kinect) sheds new light on this question. As stated by Pedersoli et al. (2014) "Kinect approach is non-intrusive; sensing is passive, silent, and permits to overcome the limitation of robustness, speed, and accuracy of typical image processing algorithms by combining color images and depth information." If Microsoft Kinect found a great echo in the scientific community to propose new ways of interaction, most of the articles focus on the "technical" side [e.g., (Raheja et al. 2011; Zhou et al. 2013)] or the interaction metaphors [e.g., (Song et al. 2012)]. To the best of our knowledge, few scientific studies have focused on the real contribution of this emerging technology through an extensive study (with qualitative and quantitative assessment) comparing a complete system to a common hardware.

The objective of this paper was to report the design and evaluation of a solution to capture hand movements without sensors, making real-time 3D seamless interaction possible for navigation and manipulation tasks. We mean a solution as a combination of a technology, computer vision algorithms and appropriate interaction modalities. Indeed, the work on modalities of interaction is crucial because users are manipulating a virtual object and laws of physics do not apply in this case (Poor et al. 2013). We need to provide users simple metaphors to allow them to understand the interaction means and predict the result of their gestures. Contrary to other approaches that aim to associate classic interfaces with seamless systems (Wang et al. 2011), and according to recent research works [e.g., (Song et al. 2014; Raheja et al. 2011; Pedersoli et al. 2014; Rodríguez et al. 2013)] and more precisely those that focus on the capture of the state of the hands (open or close) [e.g., (Yeo et al. 2013; Song et al. 2012; Jaehong et al. 2013; Unseok and Tanaka 2012)], our objective was to propose a fully natural interaction. Our originality lies in providing a complete system and to assess its interest with a detailed experimental study. The intended use concerns desktop applications and, more specifically, large-scale immersive environments where the user can stand and benefit from stereoscopic vision. Application domains are numerous and include health care, including rehabilitation (Movea 2009; Zhou and Hu 2008).

After presenting an overview of the current means of interaction, we will detail the design of our solution and the proposed interactions modalities. We have focused on three tasks that are the most common when interacting in virtual environments (Coquillart et al. 2003): object selection, object manipulation, and navigation. We will then present our experiment conducted with 20 participants, comparing our solution and data gloves paired with magnetic sensors, relatively to performances, acceptability, and preferences of participants. We will end this paper with conclusions and perspectives.

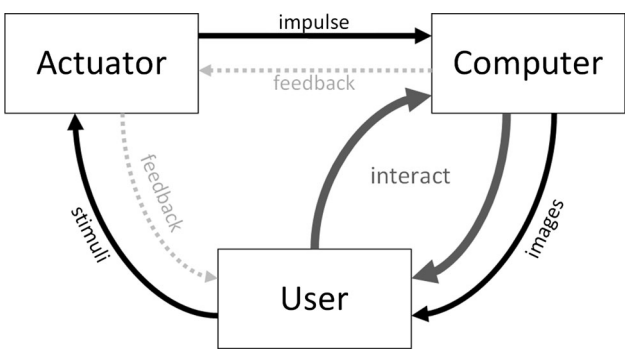


Fig. 1 Model of human–computer interaction (*left*), from (Norman 1988). Our view of the Norman's model for human–computer interaction (*right*)

1 Related work

The capture of hand movements can rely on "software" or "hardware" techniques. Software techniques imply that a physical interface captures a video or a 3D data stream, but the core work is based on image processing. "Hardware" techniques can be of various kinds: electromagnetic, mechanical, optical, ultrasonic, etc. Each solution, software or hardware, seamless or not, can be classified into three major families (Hayward and Astley 1996): low (2–3 DoF), high (4–6 DoF), or very high (more than 6 DoF) degree-of-freedom device. Hardware systems are the most commonly used, because they are widespread and allow a high accuracy and a high reliability of data, whatever the technology used: mechanical exoskeletons, date gloves (pressure sensor, optical fiber, bending sensors), optical systems (based on passive or active markers), and magnetic systems. These systems have common drawbacks: they are expensive, require a calibration phase, and need sensors or equipment on the hand(s) of the user. For example, wearing a glove can be uncomfortable during long work sessions

(Wang et al. 2011). For the design of our solution, such devices do not meet the constraint of transparency so we focused our study on software systems.

1.1 Software systems

Much research work has been focused on the recognition and tracking by one camera (monoscopic systems) or more (stereoscopic or multiview systems) by means of image processing rather than the usage of captors. This area is called computer vision and processes information in both 2D and 3D. Computer vision usually requires a combination of low-level algorithms to improve the quality of the image and high-level ones to "understand" the picture (recognize patterns and interpret them). Each technique is presented in Appendix 1.

1.1.1 Techniques and algorithms used in monoscopic vision

To ensure accurate tracking of hand movements in three dimensions, monoscopic vision is often associated with other techniques, which can be based on the use of markers or not.

Approaches with markers Color markers or patterns (binary images) are placed on each of the user's fingers (Pamplona et al. 2008; Hürst and Wezel 2013). Colored gloves (Dorner 1994; Geebelen et al. 2010; Tokatli 2005; Wang 2011) and colored rings solutions (Mistry et al. 2009) are based on the color, which becomes the essential information required to follow the hand or finger movements.

Approaches without markers Some systems are based on the recognition of skin color (Shen et al. 2011); they only obtain good results if the brightness is constant (Hassanpour et al. 2008). The "template-matching" technique is used in many studies to detect and track hand movements and is based either on the contours of the hand (Mohr and Zachmann 2009; Stenger 2006) or on silhouettes of the hand (Mohr and Zachmann 2010a, b) or color (Stenger et al. 2006). Contours are more characteristic for articulated objects but they can be difficult to obtain because of external constraints such as lighting or camera settings. The recognition of silhouettes uses the silhouette (Prisacariu and Reid 2011; Tosas 2006; Tosas and Bai 2007) of the user's hand to position and best match a 3D model. This method requires many templates for a single match, which has a significant impact on the computation time and thus on the real-time aspect.

Finally, systems based on 3D models re-adjust an articulated 3D model of the hand by adapting the most probable posture corresponding to the image seen by the camera (Ouhaddi and Horain 1998). This is one of the most widely used techniques to estimate the postures of the hand from a single video stream but it does not allow the user to know the position of the model in 3D.

Monoscopic vision techniques are not enough to obtain accurate 3D information and track the movements of the hand in three dimensions. These techniques have to be coupled with other technologies (e.g., sensors positions) or adapted to stereoscopic systems to allow 3D interaction in virtual environments.

1.1.2 Main algorithms used in stereoscopic vision

Two or more cameras can provide depth information through stereoscopy. All systems seen above can therefore be reused in stereoscopic systems [e.g., systems based on skin color (Elmezain et al. 2008) or on colored gloves (Theobalt et al. 2004)]. Systems based on 3D models are also suitable but the cloud of points coming from the depth map replace former 2D images (Dewaele et al. 2004). There are also multi-view reconstructions (Hong and Woo 2006) that require multiple cameras around the hand and that are based on different methods like "shape from silhouettes" (Ueda 2003). Finally, (Schlattmann and Klein 2009) suggest a system, based on the usage of 3 cameras and a technique called "pose estimation" that allows the user to manipulate the information in 3D.

To allow interactions with the 3D objects, the system detects the posture of the hands thanks to a coarse model obtained from the three cameras. Some systems use only 2 cameras, for example those that are for CAD applications (Wang et al. 2011) but are limited to some gestures (i.e., pinching). This system is inexpensive and allows bimanual interaction, but the user must keep a particular pose of the hand for a moment to allow the detection of the action, which is unsuitable for real-time and direct interaction/manipulation.

1.1.3 Conclusion

Algorithms used in monoscopic vision do not allow real-time 3D interaction. Stereoscopic solutions are more effective (Stefano et al. 2004) but they have to be improved to allow real-time processing with sufficient resolution. In addition, methods used to obtain the third dimension require a long computation time that leaves few resources available for additional treatments (i.e., capturing and tracking the movements) and make the final system too slow to be considered real time. Finally, stereoscopy brings about other complex difficulties such as parallax and lens distortion.

1.2 Emerging technologies: 3D cameras

No hardware solution is directly suitable for all possible virtual reality applications. The choice of a system over another requires a set of criteria and constraints related to the tasks to be performed. These constraints are numerous: real-time capture, nature of the movements to track, the absence of physical connection, size of the workspace, accuracy, resolution, environment, or price. Software solutions can override most of these constraints to propose seamless low-cost systems but algorithms need to be strictly optimized to ensure real-time interaction. An interesting solution would be to release the computer from depth computation by directly moving them onto the capture system.

Such systems already exist and are known as 3D cameras (Lange 2000; Lange and Seitz 2000). An evaluation of the use of these cameras in the field of computer graphics was done by (Kolb et al. 2009). Major techniques are triangulation-based cameras (May et al. 2007) and "time-of-flight cameras" (TOF), also called RGB-D cameras (RGB-Depth) or Z-cam. Microsoft's Kinect uses the first technology. The Kinect gives 3D depth and relies on computer vision algorithms to detect objects or persons. Thanks to the OpenNI library, it is possible to determine the positions of the user's limbs, and relative to the scope of our study, this technology allows hand tracking without additional sensors, contrary to data gloves that require electromagnetic sensors, for example. Other advantages are its relative insensitivity to ambient light (the system is based on infrared light and able to work both day and night) and to magnetic disturbances. Based only on distances (with the depth map) and not on color, it is possible to isolate the person facing the screen and to focus on his hands, removing people and objects in the background. 3D cameras have the potential to overcome most of the limitations of other devices (e.g., colors, lighting conditions, metal sensitivity, and use of sensors). Moreover, the main advantage of a Kinect-like camera is the price/quality ratio and the OpenNI library, which is freely available (Microsoft's Software Development Kit (SDK) was not released at the moment of our developments). On the other hand, it is important to mention that the Kinect also has some drawbacks. First, its framerate (30 fps) could appear low compared to cameras used in optical tracking systems (up to 1,000 fps), even if this is enough for real-time interactions. The Kinect sensitivity to infrared light might be a problem if used with an optical tracking device or when there are some reflective or transparent objects. When infrared dots (from the light pattern) hit a reflective object, light is deflected and the Kinect cannot provide any depth information for these points. Finally, if an object is close to the camera, a considerable shadow is present, due to the distance between the infrared projector and the depth camera.

1.3 Our view of Norman's model

The use of a 3D camera allows us to eliminate the feedback between the user and the device, and between the computer and the device, making the system completely seamless to the user. Figure 1 shows our view of Norman's model presented above: the large gray loop corresponds to what the user perceives and the black loop corresponds to what is really happening. When the user moves his hands, his gesture is visually relayed to him by the computer: the visual feedback associated with the absence of wearable sensors makes the device totally seamless, even if the device is still present. Computations must be performed in real time and with the lowest latency between the real action and the visual feedback returned by the computer.

2 Method

We took an experimental approach based on a comparative study between our solution and a common and functionally equivalent system. We focused on three common tasks when interacting in virtual environments (Coquillart et al. 2003): object selection, object manipulation (orientation, position, rescale), and navigation. The system that we compare to ours is composed of data gloves and magnetic sensors. We chose them because they are commonly used in businesses as well as in research. They are generally cheaper than optical systems and therefore closer to a low-cost system, such as the one we propose.

2.1 Design of our solution

2.1.1 Implementation

For our solution, we chose Microsoft's Kinect 3D camera, as it offers 30 fps and a resolution of 640 × 480 pixels, far superior to other cameras (e.g., Mesa SR4000, PMD). The user sits at a table facing a screen with a Kinect on it. This way, the Kinect can retrieve the position of the user's hands without disturbing the field of view or the workspace (150 × 150 × 100 cm) of the user (Fig. 2, acquisition unit). All the objects to manipulate are virtually placed between the Kinect and the user. The user's movements in the real world correspond to the same as in the virtual one; we work on a 1:1 scale. The distances are relative to the position of the Kinect (or the antenna for the second system).

To determine the 3D position of the user's hands in the workspace and the status (whether they are opened or closed) in real time, we used the OpenNI SDK

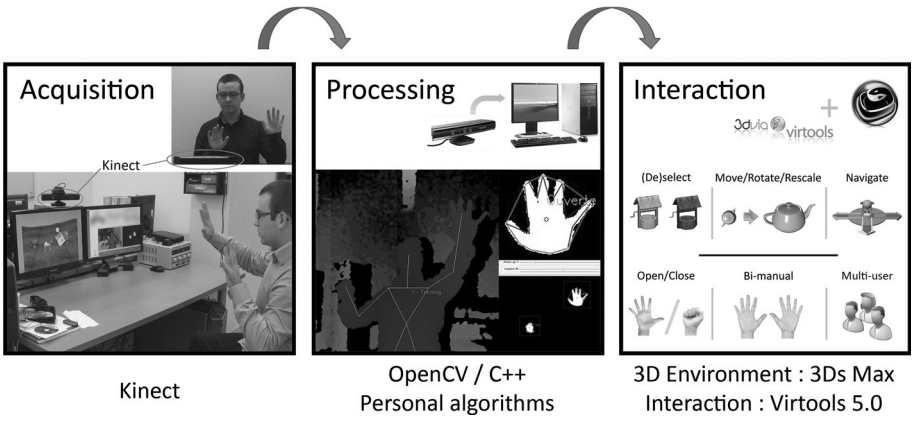
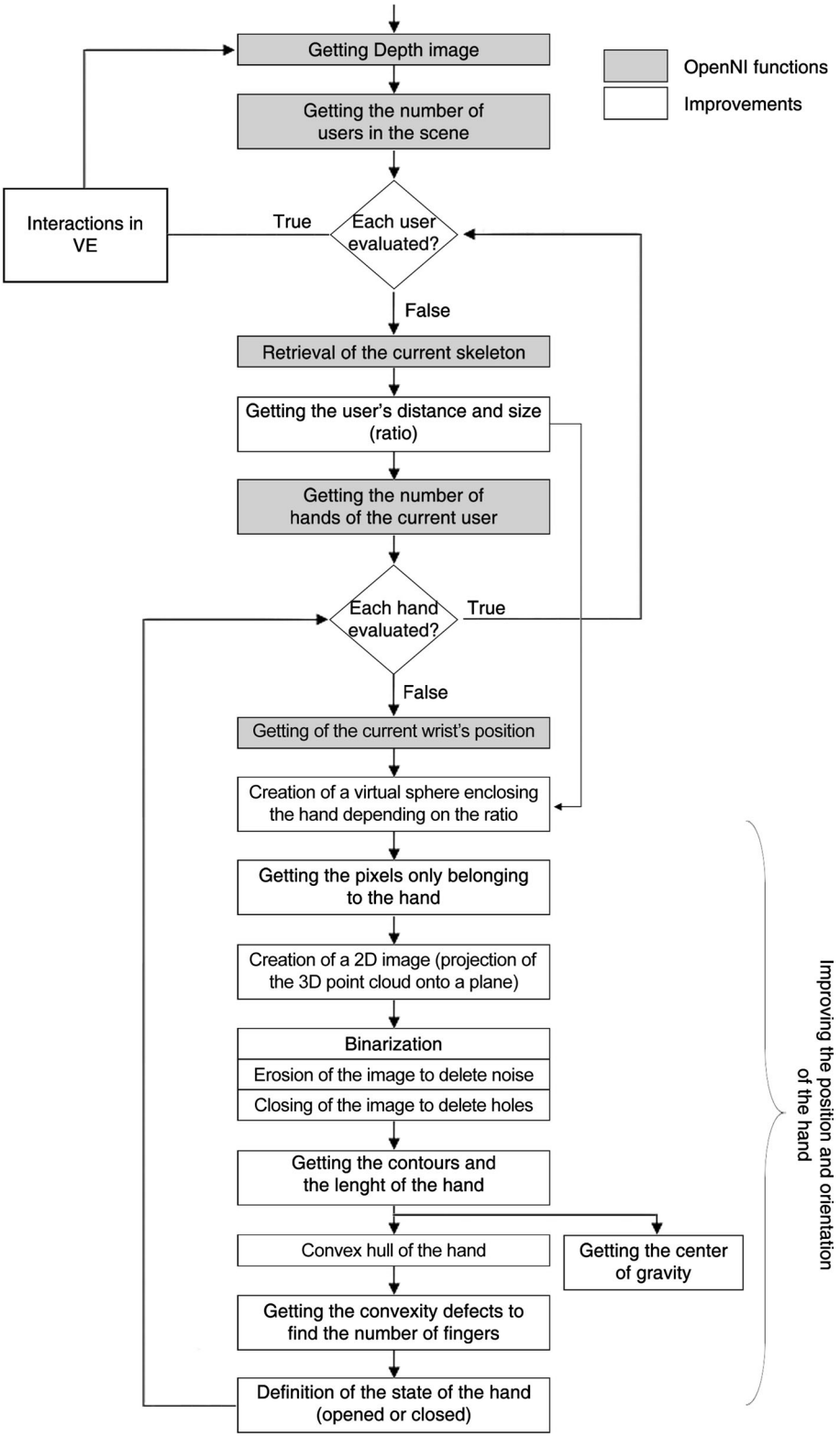


Fig. 2 How the solution works and what it can do

(Primesense), in order to get the 3D video stream (actually 2, 5D), to apply a skeleton to the user and to recover the positions of his hands. We were also able to differentiate the left hand from the right in order to take into account the lateralization of the user when handling objects. However, the SDK gives no information about the orientation or state of the hands. We thus developed our own algorithms in C++ to determine the status of the user's hands, using the OpenCV library. At the time of development, the OpenNI SDK only allowed the user to know the exact position of the wrist in 3D, but it was not possible to get the exact positions of the center of the palms nor to know whether the hands were currently opened or closed. In order to interact accurately in 3D with virtual objects, it was then necessary to have more information and to develop new algorithms.

The complete processing is described as follows (Figs. 2, 3). Once the depth image is acquired by the Kinect (1), it is possible to determine how many users are present in the scene (2). For each of these users, a skeleton is retrieved (3) and it is also possible to get the position of the wrist (4). Steps 1–4 are available directly through OpenNI functions (gray rectangles in Fig. 3); all the following steps compose the algorithm we developed to improve the accuracy of the tracking and to determine the state of the hands. When the skeleton of the current user is acquired, the distance and the size of the user are recorded to get a ratio (5) which will be used latter. The distance is obtained by simply retrieving the position (x, y, z) of the belly of the user. As the reference (0, 0, 0) is in the center of the Kinect, the distance of the person is obtained directly. To get the size of the person, we calculate the distance between the highest point of the head and the ground. Then, for each hands of the user, using the position of the wrist determined in step 4, we create a virtual sphere that encloses the hand (6). The radius of this sphere depends on the ratio calculated before. This fictitious sphere allows to only keep the pixels belonging to the hand and that must be taken into account (7) for the following steps, all the others pixels are not used. To know whether a pixel belongs to the sphere, it is necessary to compute for each pixel the Euclidian distance between this pixel and the wrist. If the distance is less than the radius of the sphere, the point belongs to it and it is stored in a structure. This allows to reduce the number of computations in the following steps and to improve the global speed of the algorithm. The following steps are done on 2D images representing the projection of the 3D point cloud onto the plane of the camera (8). A binarization is done in order to have a black and white image (9), white pixels represent the hand of the user and black pixels are considered as background. The binarization process was used in order to reduce the number of computations as in this case there only is one channel to compute instead of three (red, blue, green). A first filter—a simple erosion operation done through a convolution mask—is applied on the image in order to delete the noise around the hands (10). Then, morphological "closing" is applied on the image (11). These steps are necessary because sometimes some holes can appear in the hand, resulting from the binarization, the erosion, or from reflective object like jewelry or rings on the fingers. Once these steps are done, the image is improved and the hand is clean from noise or holes, the contours and the length of the hand are then retrieved (12). It is also possible to calculate the center of gravity of the shape by calculating the mean position of all the points of the hand, this mean position represents the center of the palm (13). The 3D coordinates of this point are then retrieved from the depth image. The position of the hand is thus accurately determined (14). To detect whether the hand is opened or closed, the method of (Homma and Takenaka 1985) was used. The convex hull of the contour of the hand is computed, and then the convexity defects are determined thanks to an OpenCV function (15). If the defect is

Fig. 3 Process diagram



longer than 2 cm, the finger is considered as raised. It is then possible to count how many fingers are down, if two or more fingers are raised, the hand is considered as opened, otherwise it is considered as closed (16). To have an idea of the orientation of the hand, points all around the center of the palm are taken. The 3D coordinates of these points give a circle from which the orientation of the hand can be determined (17).

These different steps are done for each hand of each user present in the scene. Once the accurate position and the state of the hand are computed, interaction with virtual objects can be done. Steps 5–17 are improvements that OpenNI does not offer yet (white rectangles in Fig. 3). All these steps require 23 ms from the acquisition of the depth image to the determination of the position and state of each hand.

Our algorithm (see Fig. 3) is thus able to recognize the open or closed status of each users' hand in front of the Kinect (up to 3 simultaneous users reliably). Although entirely possible, this study is not concerned with this aspect.

Based on the work of Kim et al. (2014), we measure the latency of our system: the latency was measured by moving the hand and then abruptly stopping, while a video of the hand and the application was recorded using a high-speed camera (120 fps). Analyzing the video, frame by frame, gives us a mean latency of 287 ms for ten attempts (SD 22 ms). This value corresponds to the global latency of the system (from the data acquirement to the visual feedback on the computer screen). In video games (running at 30 fps), the mean latency is about 132 ms. The latency induced just by the data acquisition of the Kinect is about 90 ms, and Livingston et al. (2012) measured the latency of the skeleton tracking, which is about 146 ms (for one skeleton in front of the Kinect). Our results are consistent with other works, like for example with the work of Kim et al. (2014) where Kinect tracking adds between, 200 and 400 ms compared to a robot master. Even not negligible, the latency is enough low to not disrupt the user experience.

The virtual environment (VE) has been modeled in 3DS Max, and interactions have been developed in Virtools 5.0. Interactions we developed are selection/deselection, manipulation (moving, rotation, and resizing) of objects, navigation in the VE, and control of the application. For this, we have developed new ways of interaction that we wanted to be intuitive and close to our real actions so that anyone can use our solution without special knowledge in computers or virtual reality.

2.1.2 *Interaction modalities*

Besides simple and intuitive ways of interaction, we added visual feedbacks to indicate the user's hands position (2 small spheres) as well as their status (green for opened/red for closed) as it is suggested in (Mason and Bernardin 2009). Several actions only require the use of the dominant hand (initialized for each user): selecting/unselecting an object, displaying the menu, moving objects, and navigating in the virtual environment. We considered the lateralization of the user in order to make our solution easier and to enable better accuracy as the predominant hand is capable of producing fine-grained gestures (Guiard 1987). More complex tasks like resizing and rotation require the use of both hands.

Interaction techniques have aroused great interest in the scientific community (Beaudouin-Lafon 2004; Klein et al. 2012) but they are mainly dedicated to 2D applications. Today, manipulation of objects in 3D space becomes accessible to the general public mainly through video games devices (e.g., Nintendo Wiimote, Sony PS Move). Like (Hand 1997; Laurel 1986), we wanted to develop interaction modalities that are as close as possible to real actions. Many works deal with 3D interaction techniques for seamless devices (Song et al. 2014; Lin et al. 2013; Song et al. 2012) but there is still no universal solution adopted by everybody. We based our interaction techniques on different works like the ones of Fiorentino et al. (2013) for the scaling modality for example. Concerning the rotation, different modalities have been proposed like the one of (Song et al. 2014) where the rotation is intuitive but need the use of an eye-tracker to select the center of rotation. Other works like (Song et al. 2012) or (Soh et al. 2013) deal with the rotation modality but these techniques are not intuitive, we nevertheless used the idea of (Song et al. 2012) concerning the possibility of doing the movement in several times, they called this technique the "pedaling" motion.

As (Zhang et al. 2014), we can distinguish two kinds of gestures: gestures which trigger a control (called *offline*) and those which are interpreted and processed in real time, like object manipulation (called *online*). Regarding our system, we could make the distinction between movements that are used to validate an action (open hand or closed hand) and those directly related to the manipulation and the navigation in virtual environment. This is consistent with (Bowman and Hodges 1997) by considering two steps in object manipulation: grabbing and manipulation interaction (orientation and position of the object). Consistent with the recommendations of (Bowman and Hodges 1997) to separate grabbing and manipulation to ensure good overall usability, we implemented these two steps separately. We compared the most common techniques presented in the literature to our constraints namely that the objects to select/manipulate directly in the workspace of the user, that there is no occlusion between the objects, and finally, that the navigation is done between two scenes along a path. These tasks were chosen deliberately simple to validate our system.

Poupyrev et al. (1998) proposed to categorize selection techniques among exocentric and egocentric metaphors. In the latter, they distinguish two metaphors, "virtual hand" (such as arm-extension techniques) and "virtual pointer" (mainly ray-casting techniques). If ray-casting techniques

(Mine 1995) are convenient to reach an object, they make the manipulation complex due to the use of a global coordinate system (especially for rotations). Moreover, the selection of small or distant objects through virtual pointing remains to be a difficult task (Argelaguet and Andujar 2013). On the other hand, arm-extension techniques are convenient to manipulate objects directly in their own coordinate system. That is why we have privileged direct manipulation techniques for interacting with objects. Indeed, until the arrival of Kinect, data gloves were naturally used for "virtual hand"-based interaction (Ughini et al. 2006); our system is proposing to replace the data gloves; we focused on that kind of metaphors which allow a better immersion as the user can see a representation of his hands. For selection, as all items are available in the workspace of the user, we have not had to implement methods like go–go (Poupyrev et al. 1996) for example. The ray-casting is advantageously replaced by the selection method which is called 3D paint-to-select by (Zhang et al. 2014). The constraints (no occlusion, limited workspace, and limited precision) of our system allowed us to avoid the use of interaction metaphors such as world in miniature (Stoakley et al. 1995). Our approach is therefore closer to hybrid techniques such HOMER (Bowman and Hodges 1997), as we separate selection and manipulation. But we used a simpler technique than ray-casting for selection and we implemented direct manipulation rather than arm-extension techniques. In this manner, users can grab and manipulate objects simply using natural motions (Robinett and Holloway 1992).

Finally, new paradigms tend to adapt multi-touch gestures used in today's smartphones to manipulate 3D objects in virtual environments. For example, (Nan et al. 2013) proposed a system dedicated to image segmentation and composition in CAVE using fingers interaction. The bimanual interaction is close to our system, and resizing is implemented in a similar manner (the size of the object is linearly indexed to the distance between the hands). For cons, the implementation of the moving is more complex, because the reference of the movement is the midpoint between the two, imposing to maintain a constant distance between hands during translation, while not allowing a validation action. The rotation is not necessarily intuitive for the chosen axis. One improvement proposed by (Zhang et al. 2014) allows for 3 handling tasks simultaneously, but requires the use of a wand-type device (a button allows to switch between interaction mode and pause mode).

Simple tasks For object selection, the user must move his dominant hand on the object and close his hand on it for 2 s. When the object is selected, its color turns blue. To unselect, the dominant hand must be held opened on the object for 2 s, when it is deselected the object goes back to its original color. To move a selected object, the user has to put his dominant hand (closed) on the object. The object is thus attached to the hand and then follows its movements while it remains closed. When the user opens his hand, the object is released. Our implementation of the selection follows the guidelines of (Bowman et al. 1999) as the user can see the collision between the sphere representing his hand and the object to select (object indication), close his hand to select an object (confirmation selection) and see the changing of the color of the object when selected (visual feedback).

For the navigation, we chose the metaphor of a joystick. Navigation has been described as consisting of 2 components: travel (the task of moving from one location to another) and wayfinding (the task of acquiring and using spatial knowledge) (Bowman et al. 2004). Interaction techniques for travel [e.g., (Mine 1995; Bowman et al. 1997)] and wayfinding aids [e.g., (Darken and Sibert 1996)] have been proposed in the literature. Among three general interaction paradigms for 3D virtual environments proposed by (Ware and Osborne 1990) (eyeball in hand, scene in hand, and flying vehicle control), we chose an egocentric technique: flying vehicle because it is more consistent for our system where the user moves from one selection to another scene. Here, there was no way to inspect an item as it can be possible with techniques such as ViewCube (Khan et al. 2008). To get as close as possible to free navigation, and given the simplicity of the navigation task we implemented, we excluded "point-of-interest" (POI) techniques, although they may have interests such as quick navigation (Haik et al. 2002). Moreover, in the context of the discovery of an environment, it is not relevant to guide the user. Recent POI techniques as "Drag'n go" (Moerman et al. 2012) are often more suited to the use of a device (as the user must "validate" the target he wants to reach), and ask an heavier cognitive load. Finally, our environment was not specific and did not require a variable level of details, which is why we have not adopted any multiscale techniques like the one proposed by (McCrae et al. 2009). Concerning the wayfinding, many arrows are placed in the environment to indicate the direction to follow. Regarding the travel, it appeared that we needed a starting point for the user and a navigation direction. We wanted to have a visual reference for the user to know at all times how he moves in space relative to his starting point. The solution of a virtual joystick was chosen because the rest position of the stick of the joystick is used as a reference and the user can easily see in which direction he moves. The metaphor of the joystick allows the user to finely control the speed of movement at every moment, navigating, as speed of motion is important to effectively navigate in 3D environments (McCrae et al. 2009). For people who never used a joystick, it remains intuitive because they just have to move the hand in the direction

they want to go from a starting position. To navigate, the user closes his hand on the virtual joystick for 2 s to select it and moves his hand in the direction he wants to go. When the joystick is selected, a frame and arrows appear to show the user the correct direction as well as the final location.

Complex tasks The following tasks are more complex because they require the coordination of both hands. By using the analogy of a spring that can be compressed or stretched to change its size, users have to spread or bring their hands closer to enlarge and reduce the object. Resizing is homogeneous in all three axes. This solution is close to solutions proposed recently [e.g., (Fiorentino et al. 2013; Song et al. 2012)].

For the rotation, the user must put his dominant hand (closed) on the object. The representation of the hand disappears and is automatically positioned at the center of the object. By moving the secondary hand (closed) along the X axis (or Y or Z), the object rotates in the same direction. Contrary to others approaches like the use of sheet of paper (Song et al. 2012) with impose to use a tangible object, the rotation with two hands (Soh et al. 2013) which is not very intuitive, and unlike (Wang et al. 2011), we chose to simplify this interaction, even though by doing this it becomes less realistic.

2.1.3 Application control: a pie menu

To allow the user to switch from one action to another, we set up a menu. We excluded gestural language like in (Soh et al. 2013) which we considered to be too complex and which could lead to misinterpreted gestures; moreover, they oblige the user to memorize the gesture for each action he wants to do. We also excluded voice commands because we did not want to introduce new variables related to multimodality. Levesque et al. (2011) also proposed a 3D bimanual gestural interface using data gloves for 3D interaction; the left hand can select interaction modes while the right hand is for the interaction itself, like the rotation of an object. This solution was not chosen because we wanted the user can use his two hands to perform the different actions. We chose a pie menu (Fig. 4) as previous work demonstrated the importance of these menus over linear menus (Callahan et al. 1988). They allow increased speed while minimizing errors of selection. In addition, the distance between the point of activation and all the different items is the same, due to the circular organization.

In order to bring up the menu, the user closes his hand, opens it, and keeps it open for 2 s. The menu appears at the exact position where the user closed his hand. To choose an action, the user must stay on the item selected for 2 s. The method we have chosen (i.e., a waiting period of 2 s) for the validation of the selection (Gratzel et al. 2004) causes a lag in the interaction which is a disadvantage of our solution. But it was the only solution compatible with our goals of reliability, stability, and simplicity, unlike other methods, such as those outlined by (Schlattmann et al. 2009): a physical button (not seamless), head movements (unnatural), speech recognition (multimodal). Moreover, selecting the closest object (Wang et al. 2011) is not viable in complex virtual worlds. The criteria for reliability, stability, and simplicity correspond to criteria that any system should meet to be both useful and usable (Loup-Escande et al. 2011). The reliability indicates here that the proposed system allows the user to make the proposed tasks (i.e., selection, navigation, and rotation). Stability is related to the use of capture device: visual feedback of hands avatars should faithfully follow the gestures of the user. Finally, the simplicity corresponds to the simplicity in the handling of the system (before the interaction) and then in the understanding of the different features, more specifically the interaction metaphors. Even if new devices such as Leap Motion (LeapMotion 2012) still have some limitations and need to be improved in tracking fingers, it will certainly allow the user to more accurately know the position of the fingers and thus recognize more complex gestures in real time, a feature which is not possible with a Kinect.

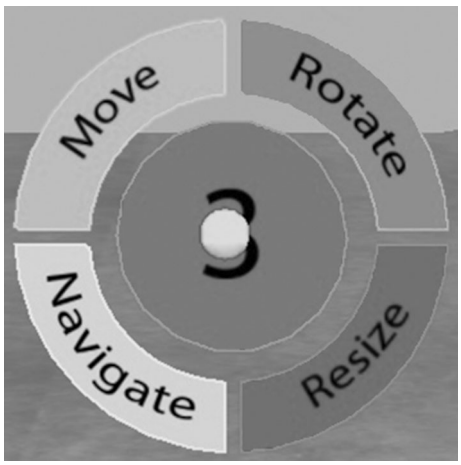


Fig. 4 Pie menu with four choices

2.2 Tasks, participants, and procedure

2.2.1 Virtual environments tasks: design

Manipulation tasks are divided into three stages of gradual difficulty. The three objects of the first scene are only associated with a single interaction, whereas those of the second scene are associated with two basic types of actions (e.g., moving and rotating). The last scene contains only one object to be moved, rotated and resized. To move from one scene to another, the participant must "navigate." The

Table 1 Sequence of tasks (*M* moving, *S* scaling, *R* rotation, Nav. = navigation)

Scene 1									
Task1			Task 2			Task 3			Task 4
Well			*Horse*			*House*			
M	S	R	M	S	R	M	S	R	Nav.

Scene 2										Scene 3		
Task 5			Task 6			Task 7			Task 8	Task 9		
Globe			*Clock*			*Computer*				*Boat*		
M	S	R	M	S	R	M	S	R	Nav.	M	S	R

Gray cells indicate actions not allowed

set of subtasks is summarized in Table 1. The tasks consist of matching objects with a transparent model. The game is displayed in perspective, allowing the user to better evaluate the distances between and the orientation of objects. Items are in the same conditions (position, orientation, scale) in both devices in order to obtain comparable data from both systems. Learning effects are controlled by mechanisms that ensure internal validity of our experiment (see Sect. 2.2.5). Figure 5 shows three objects in a scene before and after manipulation.

2.2.2 Participants

The experiment involved 20 volunteer participants, 8 women and 12 men, aged between 21 and 34 (average = 26; SD = 3.7). All were experienced users of computers and had a college BA or BS. The majority were graduate students in virtual reality, or members of a Virtual Reality Laboratory (PhDs, Research Engineers). The experiment was presented as a comparison between two VR systems without any mention of "our" solution. Participants were "naïve" as the experiment took part at the beginning of the academic year: they knew neither our work nor the person performing the experiment (who had no link with them). As the number of participants was relatively small, we were careful to restrict the age range to limit inter-individual variability with respect to performance and subjective preferences. The average age of our relatively small group of participants is explained by the fact that we wanted to first validate qualitatively and quantitatively the interest of our solution with an audience who has a certain ease with the technology and better acceptance of change.

2.2.3 Material, procedure, and instructions

We used a computer with an Intel Core ™ i7 930, 6 GB of RAM, an NVidia GeForce GTX480 graphics card, and a 22″ 120 Hz screen. For the "3DCam" solution, the Kinect is placed above the screen (16″ from the table) and interferes neither with the vision of the user nor his workspace. In the case of "3DGloves," a magnetic sensor (a Polhemus Patriot) was attached to each glove (5DT glove) and all were connected by wires to a computer (Fig. 6).

The objective of each task was the same: matching precisely the object and its model. We did not ask participants to accord more importance to precision than to the time necessary to perform the 9 tasks.

The experiment was divided into different stages: after completing an identification questionnaire, the participant received instructions explaining the experiment, the tasks, and the proposed interactions. These instructions were also given in a written form and freely available during the experimentation. The participant began the experiment by a learning process (unlimited time). Once the user was ready, he carried out the 9 tasks for which data were recorded. After that, he filled in a final questionnaire concerning his or her feelings, comments, and subjective judgments.

2.2.4 Working hypotheses and measures

We formulated four hypotheses concerning our comparative study: a 3DCam solution provides greater accuracy in manipulation tasks due to the absence of equipment to wear (H1), better execution times (H2), a higher level of overall acceptability due to a better comfort, a better efficiency and

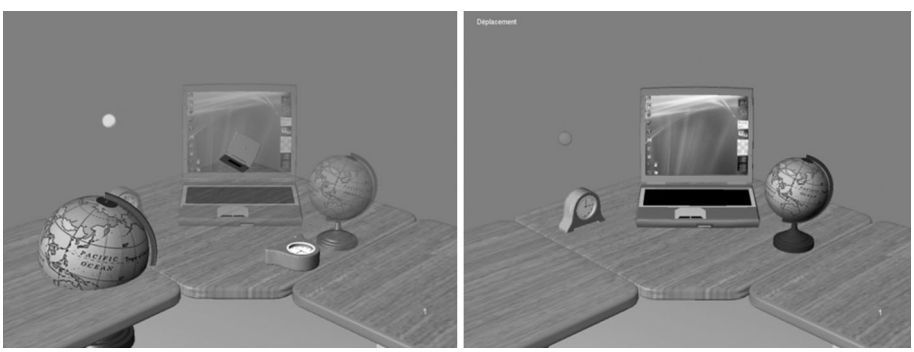


Fig. 5 Scene 2 with 3 objects (*globe*, *clock*, and *computer*) before and after manipulation

Fig. 6 Participant with the "3DGloves" system (*left*) and the "3DCam" system (*right*)

a better effectiveness (H3), and obtains participants' preferences due to a higher feeling of immersion as well as a greater ease of use (H4) compared to the 3DGloves system. These hypotheses, if confirmed, will allow us to come to the conclusion that the 3DCam is superior to the 3DGloves when it comes to manipulation tasks in VE associated with our interaction modalities. To study the participants' performances, specific metrics were taken with regard to each system:

- Total execution and manipulation time (seconds) of the 9 tasks and of each task: manipulation time = execution time − choices in the menu or "inactivity" periods
- Accuracy error, averaged over the three axes (for general errors), or calculated for each axis (when comparing rotation and moving errors along each axis).
    - The rescaling error is the percentage of difference from the reference scale;
    - The rotation error is a percentage calculated from the angular shift between manipulated and reference objects (error of 100 % = angular shift of 180°);
    - The error of movement is a percentage calculated (independently for each axis) relative to the reference position, standardized according the object size. A null accuracy corresponds to a shift at least the size of the object.

The study of subjective preferences and participants' comments is divided into two parts: judgments about the interaction modalities and an evaluation of each system separately and then the two compared to each other. Table 2 reviews the variables studied and the method of measurement. Responses to Likert scales are encoded and treated as numeric variables (1 = worst, 5 = best). Participants could also provide feedback through open-ended questions for each modality.

We decided to study the different criteria of the acceptability because they are clearly explained by numerous reference papers in the field of ergonomics and interaction with complex systems (Tricot et al. 2003). This explains our choice not to use standardized questionnaires such as the NASA TLX for example.

2.2.5 *Internal and external validity*

To ensure the internal validity of this experimentation, we restricted the origin and the age of the participants. We chose to rely on a population with prior knowledge in VR to avoid the novelty effect induced by the wearing of unfamiliar material which might result in a possible bias regarding the acceptability of the system. The identification questionnaire showed that participants had no particular experience of natural user interfaces (UI). However, we must mention that restricting the profile of our participants implies limiting the generalization of the possible results. We carefully prepared the experimental protocol to avoid bias on the perception of "expectations" of the experiment. It was presented as a "simple" comparison of two systems, without any references on the work done on our solution.

We also attempted to control extraneous variables from the real environment (lighting, noise, and temperature): the experiment took place in a single room, with air conditioning and no windows. We counterbalanced the presentation in order to compensate the potential learning effect which could have led to a potential variability of the results during the experimentation (improvement or deterioration). Participants were randomly divided into two groups (G1 and G2), each one with 4 women and 6 men.

2.2.6 *Statistical analysis*

To check the normality of all the distributions of values, we performed Shapiro–Wilk test since it is more powerful in detecting normality for samples sizes up to 2,000. For the few variables which did not follow a normal distribution, we chose nonparametric tests, with the exception of the study of the interaction of several factors for which we performed an ANOVA, given its robustness for type 1 errors (Winer 1971).

To analyze the influence of the system on participants' performances, we used the Student's t test for paired samples or the Wilcoxon test (depending on the distributions). Analysis of the effects of gender is based on a mixed ANOVA with one within-subject variable (the system) and

Table 2 Variables studied and measures

Studied variables	Acceptability	Global preferences	Evaluation of the interaction modalities
	Assessment of each system	*Comparative evaluation of the systems*	
Choices	Likert scales (5 modalities)	3DGloves–3DCam–Similar	Likert scales
Criteria	*Comfort*	General preferences	Global effectiveness
	Handiness	Most convenient system	Quickness
	Freedom of movement	Immersion feeling	Precision
	Tiredness	Precision	Simplicity
	Comfort of use	Ease of use	Intuitiveness
	Effectiveness		
	Global effectiveness		
	Precision		
	Stability		
	Reliability		
	Efficiency		
	Ease of use		

a between-subject variable (gender). The study of simple effects was based on a Student's t test for independent variables (two conditions) or a simple ANOVA (three conditions). Post hoc LSD test of Fisher was used to study the possible main effects. Results were considered significant when $p \leq 0.05$ and as a trend when $0.05 < p \leq 0.1$.

3 Results

3.1 Participant's performance

We compared global results of the participants for each system (for all nine tasks) and assessed the possible influence of gender on them. Then, we focused on each task separately. We concluded the presentation of the results by a further study, comparing the accuracy along the axes for rotation and moving.

3.1.1 Global results

Table 3 shows the comparison of the experimentation time (ExpT) and the total manipulation time (TMT) between the two systems. We did not observe any significant difference between the systems for ExpT and TMT. A mixed ANOVA (gender x system) on the ExpT did not show any effect of the system [$F(1, 18) = 0.772$, $p = 0.391$] or the gender [$F(1, 18) = 1.337$, $p = 0.263$] and no interaction between these two factors [$F(1, 18) = 0.157$, $p = 0.696$]. Regarding the TMT, we observed no effect of the system [$F(1, 18) = 0.05$, $p = 0.945$] or the gender [$F(1, 18 = 1.551$, $p = 0.229$] and no interaction between these two factors [$F(1, 18) = 0.261$, $p = 0.615$].

We had no significant results for the overall results or effects of sex. Total time during which participants manipulated or navigated was not significantly lower with our system, which neither allowed the validation nor the rejection of the hypothesis H2. First tests are partial because the variables are macroscopic and do not allow to highlight potential differences in execution time between the interaction modalities. These results are presented in the following section.

Table 3 Time of experimentation and the total manipulation time

	Experimentation duration (s)		Total manipulation time (s)	
	3DGloves	3DCam	3DGloves	3DCam
Mean	902.4	954.4	514.5	522.4
SD	266.7	274.0	176.6	193.5
T test				
T	1.000		0.179	
p	0.330		0.860	

3.1.2 Detailed results

For each task and each system, we studied the following variables:

- Execution time (ET) in seconds;
- Manipulation time (MT) in seconds; it is differentiated according to the moving (MTM), rotation (MTR) and scaling (MTS) as appropriate;
- The error of accuracy, given as four separate variables: moving (Err_M), rotation (Err_R), navigation (Err_N), and scaling (Err_S);

For five out of the nine tasks, we observed no significant difference between 3DGloves and 3DCam: task 1 (moving), task 3 (rotation), task 4 (navigation), task 5 (moving + scaling), task 8 (navigation). For the four remaining

Table 4 Significant differences between 3DGloves (G) and 3DCam (C)

	Task 2				Task 6				Task 7				Task 9	
	ET		MT		MTM		Err_R		MTS		MTR		Err_M	
	G	*C*	*G*	*C*	*G*	*C*	*G*	*C*	*G*	*C*	*G*	*C*	*G*	*C*
Mean	33.1	41.3	10.0	14.9	24.0	52.6	2.76	3.66	20.0	29.1	80.1	46.4	5.06	10.42
SD	10.4	17.2	3.2	8.1	6.9	30.4	1.71	2.91	11.7	18.3	49.4	32.9	3.31	10.75
Test (*T*)	2.093		2.405		4.268		2.046		1.920		3.052		2.229	
p	0.05		0.027		<0.001		0.055		0.07		0.007		0.038	

tasks, relevant data are summarized in Table 4. For the simple task of scaling (task 2, horse), ET and MT are lower with 3DGloves. For the task requiring moving and rotation (task 6, clock), we observed a lower moving time with 3DGloves and a lower error for rotation with this system. For the task requiring scaling and rotation (task 7, computer), scaling time was lower with 3DGloves but rotation time was lower with 3DCam. Finally, the task mixing the three basic actions (task 9, boat), the moving error was lower with 3DGloves.

3DGloves seem to be able to minimize the ET for the scaling tasks. This task requires less precision than the other, and the best time obtained with 3DGloves can be explained by a better recognition of the state of the hand and because the hands (i.e., magnetic sensors) do not move away from the antenna, which results in stable values.

These results refuted hypothesis H1 and did not confirm hypothesis H2. We expected a superiority of 3DCam on precision due to the absence of weight. Finally, the accuracy appears equivalent between the two systems. We can therefore say, considering current developments, that the 3DCam does not outperform a current virtual reality device (i.e., 3DGloves) in terms of performance (time, accuracy).

3.1.3 Additional study (for 3DCam only): precision along the 3 axes

The results for moving tasks (1: well, 5: globe, 6: clock, and 9: boat) are summarized in Table 5. For tasks 1 and 5, the error is significantly higher on the *Z* axis than on the *X* and *Y* axes. These findings are reversed for task 6, with a lower error on the *Z* axis. For the task 9, the error is significantly greater with the *Y* axis than the *Z* axis and the *X* axis. The results for rotation tasks (3: house, 6: clock, 7: computer, and 9: boat) are summarized in Table 6. For the task 7, the orientation error is significantly lower with the *Y* and *Z* axes than with the *X* axis. For the task 9, the orientation error is significantly lower with the *X* axis than with the axis *Y*.

Synthesis For the first tasks (well and globe), movements are significantly less accurate with the *Z* axis. For the well, this can be explained by the amplitude of the required movement associated with a problem of the recognition of the state of the hands as well as the possibility of being out of the Kinect's range. It is the same for the globe and the clock but we also observed that a perspective effect could bring about an inaccurate sensation of correct positioning. This problem was less present due to the orientation of the boat and the clock. We also noticed that results on the *X* axis are poorer when the movement is associated with a rotation. The results are less conclusive for rotation given the variability of the data (e.g., in the task 7, the rotation along the *X* axis (hand's movement in depth) is the most imprecise as it appears better for task 9). However, we thought that the rotation along the *Z* axis (vertical hand's movement) could also be less precise because this axis is the one for which we have the lowest amplitude, and it is difficult to lower the arm without moving it forward (which produces a second rotation simultaneously).

The differences are thus mainly related to the interaction and the task. For moving tasks, we can recommend finding a technique to limit the workspace and prevent the participant from being outside the scope of the Kinect. For rotations, the difficulty of separating the axes may be a justification of the variability of results (which must be confirmed by the analysis of participants' preferences and feedbacks).

3.2 Participants' preferences

We studied the participants' subjective preferences based on their answers to the final questionnaire, regarding acceptability and global preferences. We then presented the results for interaction modalities and suggestions for improvement made by the participants. We performed nonparametric Wilcoxon tests rather than Student's *t* tests as our data did not follow a normal distribution (see "Statistical analysis").

3.2.1 Acceptability

Mean scores to Likert scale for each criterion are represented on Fig. 7. The surrounded criteria indicate a

Table 5 Decomposition of the positioning error along each axis for 3DCam

Task	Task 1 (well)			Task 5 (globe)			Task 6 (clock)			Task 9 (boat)		
Axis	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*
Mean	9.62	8.70	18.58	2.20	2.30	6.84	43.51	27.74	12.75	3.21	22.98	5.07
SD	11.57	11.79	16.10	2.11	3.13	7.89	3.18	28.15	10.64	2.95	27.76	4.57
Pair	*X–Y*	*Y–Z*	*X–Z*	*X–Y*	*Y–Z*	*X–Z*	*X–Y*	*Y–Z*	*X–Z*	*X–Y*	*Y–Z*	*X–Z*
Test	1.082	4.429	4.030	0.184	2.953	3.290	2.620	2.387	12.387	3.386	3.108	2.168
p	0.293	**<.001**	**0.001**	0.856	**0.008**	**0.004**	**0.017**	**0.028**	**<.001**	**0.003**	**0.006**	**0.043**

Results in bold are statistically significant at $p \leq 0.05$

Table 6 Decomposition of the orientation error along each axis for 3DCam

Task	Task 3 (house)			Task 6 (clock)			Task 7 (computer)			Task 9 (boat)		
Axis	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*	*X*	*Y*	*Z*
Mean	1.14	2.03	1.53	3.67	3.72	3.58	2.42	0.69	0.69	1.50	3.22	2.14
SD	0.94	1.84	1.59	4.64	2.80	3.31	2.27	0.84	0.78	1.22	2.74	1.73
Pair	*X–Y*	*Y–Z*	*X–Z*	*X–Y*	*Y–Z*	*X–Z*	*X–Y*	*Y–Z*	*X–Z*	*X–Y*	*Y–Z*	*X–Z*
Test	1.946	0.978	1.266	0.054	0.182	0.109	3.562	0	3.288	3.151	1.656	1.405
p	0.067	0.340	0.221	0.958	0.857	0.914	**0.002**	1	**0.004**	**0.005**	0.114	0.176

Results in bold are statistically significant at $p \leq 0.05$

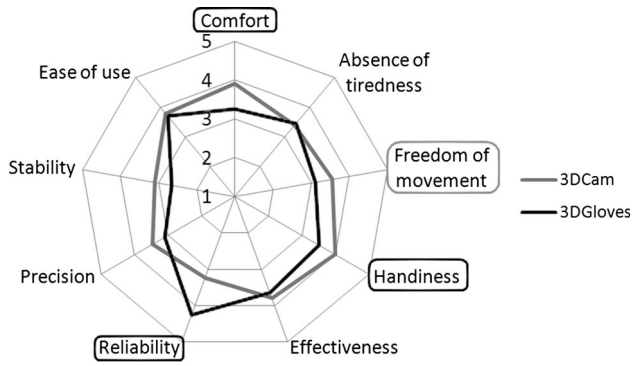


Fig. 7 Mean scores for each criterion. The surrounded criteria indicate a significant difference (in *black*) or a trend (*gray*) between 3DGloves and 3DCam

significant difference (in black) or a trend (gray) between 3DGloves and 3DCam.

There is no significant difference between the two systems concerning the tiredness ($Z = 0.302$), the precision ($Z = 1.218$), the stability ($Z = 1.510$), the effectiveness ($Z = 0.286$), and the ease of use ($Z = 0.707$).

Comfort ($Z = 2.448$; $p = 0.014$), handiness ($Z = 2.673$; $p = 0.008$) were significantly better with 3DCam; we observed a trend in favor of this solution for freedom of movement ($Z = 1.654$; $p = 0.098$). Reliability is significantly better with the system 3DGloves ($Z = 3.038$; $p = 0.002$), due to problems linked to the detection of the state of the hand by the 3DCam which was identified by several participants. Freedom of movement was considered superior to the 3DCam because participants were not wearing equipment with this solution while the 3DGloves system necessitates being connected by four cables to the computer (2 for gloves, 2 for magnetic tracking system). The handiness was judged inferior for the 3DGloves system which can be explained by the single size of the data gloves, which are not suitable for all hands. The feeling of tiredness was judged equivalent between the two systems because the weight of the gloves and magnetic sensors is negligible compared to the tiredness caused by the arm raising without support.

Acceptability appears to be better with 3DCam, which is justified by the comments and subjective judgments of the participants. The negative point, illustrated with reliability, is the bad recognition of the states of the hand that may occur sometimes.

3.2.2 Global preferences

We compared preferences according to five criteria: general preference (GP), relevance to the task (RT), feeling of immersion (FoI), precision (P), and ease of use (EoU). Participants could choose 3DGloves, 3DCam or "equal." Figure 8 shows the results.

Participants were less favorable to the 3DGloves system than to the 3DCam solution for the five criteria: for each of the five criteria, more participants preferred the 3DCam. For three out of five criteria (GP, RT, and P), more than half of the participants gave their preference to the 3DCam.

With regard to the EoU, the 3DCam seems to provide a significant advantage compared to the 3DGloves, in line with the above results (handiness and freedom of movement). Clearly, although it should be qualified with regard

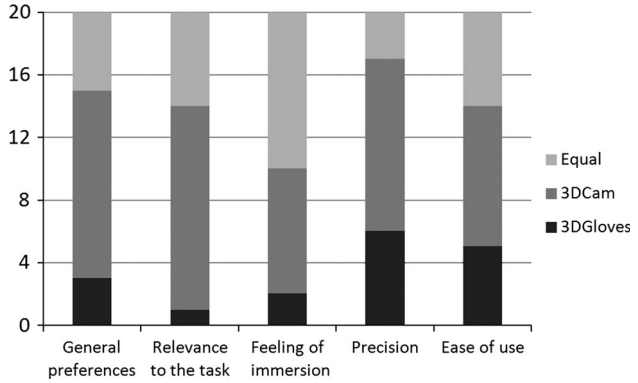


Fig. 8 Stacked histogram of the distribution of participants according to each criterion

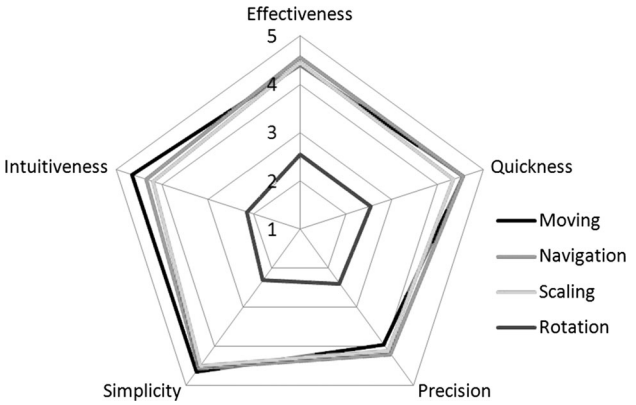


Fig. 9 Average scores for each criterion each interaction modality

to the number of participants, the 3DCam brings an equal or even greater FoI compared to 3DGloves (8 against 2, 10 without preference). We can hypothesize that the absence of equipment (and thus weight on the hands) and the freedom of movement are likely to strengthen the FoI as the interaction is the same for both systems. The short range of the magnetic tracking system (which trembles when it is too far from the antenna) was also likely to weaken the involvement of the participant in the task and thus to reduce the FoI.

Results reflect a weakness for the 3DGloves precision, which is confirmed by the participants' comments. General preferences are in favor of the 3DCam, since 12 out of the 20 would choose the 3DCam, as opposed to only 3 for 3DGloves (and 5 "equal").

3.2.3 Assessment of the interaction modalities

Figure 9 summarizes the participants' answers to Likert scale questions. It appears that rotation is significantly lower than the three other interaction modalities, for each criterion and moving is considered more intuitive than navigation ($Z = 1.897$; $p = 0.058$) and scaling ($Z = 2.310$; $p = 0.021$). Only one participant found rotation modality intuitive; for the others, two main reasons justify the weakness of rotation:

- The unusual form of the action was difficult in initial learning; this reason was mentioned by 7 participants.
- The difficulty to associate a particular movement to a specific rotation; this reason was given by 5 participants.

Moving was considered effective, simple, and intuitive by all the participants. Negative comments were more general and concerned the stability and the quality of the recognition of the hand's state. Similarly, scaling was evaluated positively by 16 participants, considered as being intuitive, fast, and accurate. Only one participant mentioned that it was difficult to know how to position both hands before rescaling. Finally, navigation was also considered very intuitive and very fast by 12 participants and as a fun experience by four participants. However, 2 participants noticed that this interaction was too sensitive and not very accurate over short distances and 5 participants mentioned excessive speed.

Moving was probably considered more intuitive than the other interaction modalities because it is the only one which is directly adapted from the real environment without an interaction metaphor (joystick for navigation, spring for scaling) or an adaptation (x, y, z decomposition for rotation).

3.2.4 Improvement suggestions

Suggestions for improvement concern comments and criticisms in general terms or specific to an interaction modality. Several solutions were proposed by participants to avoid the instability, especially with 3DGloves, which makes tasks requiring precision difficult, such as moving: modulating the speed of movement as a function of distance, or using the second hand to validate the end of a movement. These solutions could also be applied to scaling and rotation.

As far as rotation is concerned, opinions are divided. 6 participants enjoyed the simple access to the 3 axes without an additional menu but 12 participants experienced difficulties, mainly concerning the "initial learning." Seven participants suggested giving visual feedback to indicate which movement corresponds to which axis in order to help. Another participant suggested limiting the choice to a single axis and to lock the two others, which seems to go against the interest of an immediate access to the three axes. Finally, one participant mentioned rotation by a turning of the hand. Scaling was considered positively by

all participants. But some also mentioned difficulties when, for example, they started the decrease movement with hands held too close together. Since they then ended up with their hands touching each other, they were obliged to "disengage" and reposition their hands to start over. A first solution is that one hand manages the scale in one axis while the other acts like an on/off switch for the action. The second is an exponential scale when hands are close: the size of the object continues to decrease avoiding the need to release. Navigation was mainly evaluated as quick and intuitive by the participants.

These various comments and suggestions will be the basis for improving our solution.

4 Conclusions and perspectives

The work reported here is based on several premises. First, a kind of technology which aims to make the interface "disappear" to allow a "natural" user interaction is often presented as the final outcome of direct manipulation interfaces (Fuchs and Moreau 2003). These authors also suggest that motor responses must be ideally transmitted without link between man and machine. Finally, according to (Winkler et al. 2007), having no hardware to wear and making the system seamless to the user is likely to improve the sense of immersion and presence. The solution we developed tries to provide, in part, answers to these questions that are crucial in the VR field.

In this paper, we presented the design of a seamless solution dedicated to capturing hand movements for real-time interaction in 3D environments. Specifically, we were interested in the selection, manipulation, and navigation tasks. We propose a real-time solution concerning the interactions (navigation, moving, scaling, and rotation of the objects); however, calling the menu and selecting the action to perform, selecting, and unselecting objects are not done in real time because it requires poses of 2 s to validate the choices. The solution is understood here as a complete system composed of the device (Kinect), the processing and the detection algorithms as well as the developed interaction modalities. The objective was to demonstrate the value of such a seamless solution. The underlying and secondary objective was to evaluate the interaction modalities in order to propose an effective, efficient, and comfortable solution.

To accomplish this, we carried out an experiment with 20 participants and we contrasted our solution to an existing and functionally equivalent commercial system. In order to obtain comparable results, we used the same interaction modalities in the two systems. Based on our objectives, we formulated four hypotheses related to performance and subjective preferences. We expected, for our solution, a better precision (H1), better execution times (H2), a better acceptability (H3), and better subjective preferences (H4). We measured the participants' performance for nine different tasks according to global criteria or relative to each task. In addition, we assessed the subjective preferences of the participants and collected their comments and suggestions for improvements through a questionnaire.

Hypotheses H1 and H3 are linked since the absence of wearing equipment (and wired connection between the devices and the computer) should impact performances and preferences. The absence of weight and "hindrance" could help to minimize both fatigue and maximize accuracy and allow participants to experience better comfort and efficiency.

Concerning accuracy, no device performs better than the other which does not validate H1. We thought that the absence of equipment (i.e., weight) was sufficient to minimize fatigue and therefore the imprecision. Two explanations are possible: firstly, the experimentation time could be insufficient to induce fatigue and thus to affect the results; secondly, the fatigue caused by wearing the gloves and sensors could be negligible compared to the tiredness of raising an arm without support. From this viewpoint, it could be interesting to perform a longitudinal study to provide a large number of results and data related to a real use of a seamless system dedicated to interaction in VE. Longer experiment times could allow us to assess the relative importance of different criteria studied (e.g., comfort, freedom of movement…) on participants' global preferences. Contrary to hypothesis H2, execution times were no better for one device than the other. Comments from participants shed more light on these explanations since they evoked a problem of reliability in the recognition of the state of the hand with the 3DCam and a problem of trembling of the 3DGloves data when the hands were too far from the antenna. These two negative points, inherent to technical solutions, could have smoothed out the performance results. Thus, if the detection of the movements with the 3DCam and 3DGloves were the same with regard to accuracy, the reliability of the recognition of the state of the hand remains lower for the 3DCam. This problem could be overcome by adding cameras, to ensure that the hand is always correctly oriented. With a robust detection, our algorithm would be more effective. Perspectives of improvements are important: thanks to a system like the LeapMotion we probably could in the short-term consider the position of each finger with more precision, making possible a richer interaction with a much more diverse interaction language. Participant's preferences are in majority in favor of the 3DCam, mainly for the criteria of comfort, freedom of movement, and handiness. The feeling of immersion is equivalent between the two systems for

half of the participants, but is perceived as higher for the 3DCam for 8 of the 10 others. Similarly, perceived ease of use is superior to the 3DCam for nine participants, when only five evoke 3DGloves for this criterion. These results tend to validate the hypotheses H3 and H4.

Currently, our solution is able to provide an alternative to a conventional system for the main tasks usually performed in VR applications, regarding the profile of our participants. The major advantage is the cost, much cheaper than common VR equipment. The other major advantage is the potential of 3D cameras, which go far beyond the tasks we developed. Our solution can replace classic equipment while providing a better "use-value" (Loup-Escande et al. 2011). The improvements of the OpenNI or Microsoft's libraries will also enable our solution to gradually extend its potential, even if the proposed interactions can today cover the most common needs in VE.

Concerning the latency of our system, even if the value may seem significant in absolute terms, in the questionnaires filled out by the participants, no participant has been disturbed or noted the latency between their movement and the action performed on the screen. Maybe participants did not noticed this latency also because none of the interactions required to make movements needs to be fast. As moves are made at normal speed, the difference between the actual movement and the movement in the 3D environment is not really noticeable. The optimization of the code using GPGPU has the potential to bring back the latency to an effective not perceptive level. New devices like Microsoft Kinect 2 are hoped to provide shorter latencies, greater resolution, and closer range (Kim et al. 2014).

The potential applications are numerous and go far beyond only the VR field. We can imagine using gestures to control computers and navigate through applications. Uses may also be extended to health care and particularly the rehabilitation of upper limbs or in areas such as home automation where it would be possible to control different devices only with gestures (lower or raise lights, electric shutters, etc.). In art and music, it might be interesting to record the fingers movements of a virtuoso (Maes et al. 2012). These are just a few examples among many, but the range of possible applications is vast.

Ultimately, we believe that these new ways of interacting with the environment (real or virtual) will be part of our lives. It remains to improve motion capture algorithms, 3D cameras (higher resolution, higher refresh rate, improved accuracy, as promised in the Kinect2) and also improve the interaction modalities. These perspectives lead us to continue our efforts and our work in this exciting promising direction.

Appendix

See Fig. 10.

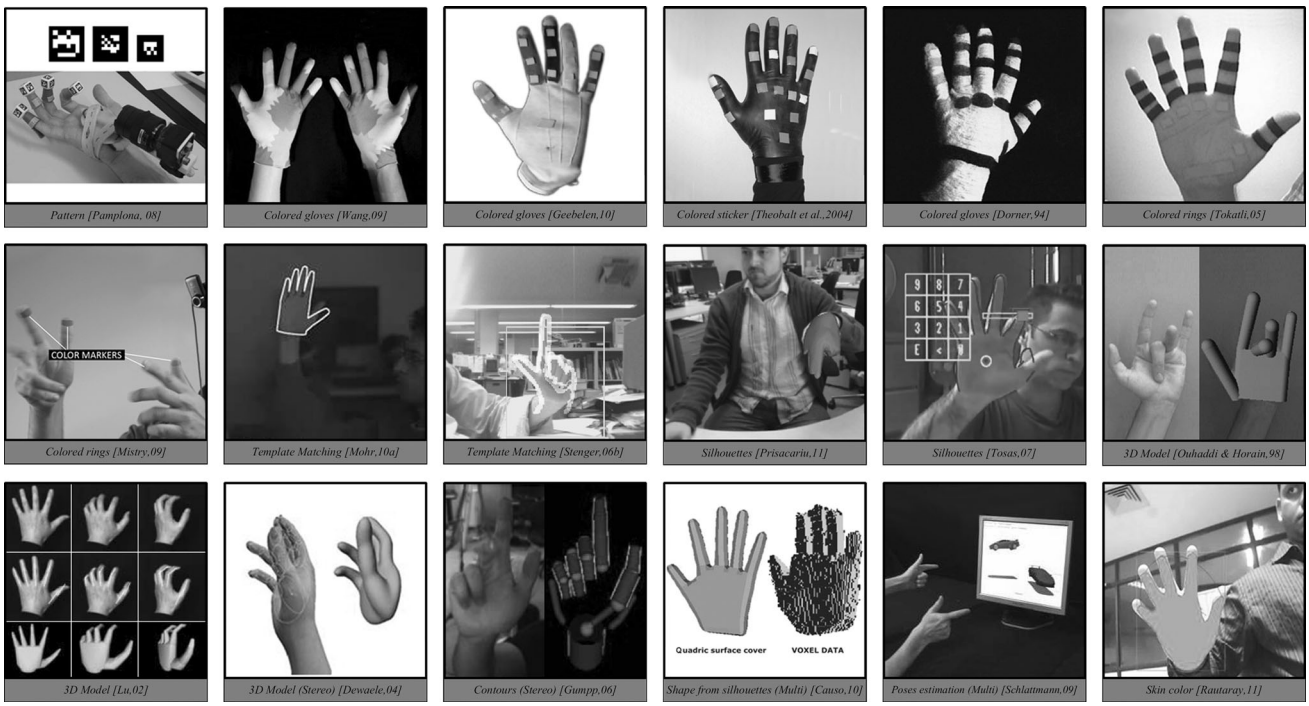


Fig. 10 Software solutions

References

Argelaguet F, Andujar C (2013) Special section on touching the 3rd dimension: a survey of 3D object selection techniques for virtual environments. Comput Graph 37(3):121–136. doi:10.1016/j.cag.2012.12.003

Beaudouin-Lafon M (2004) Designing interaction, not interfaces. In: Working conference on advanced visual interfaces, Gallipoli, Italy, pp 15–22

Berard F, Ip J, Benovoy M, El-Shimy D, Blum JR, Cooperstock JR (2009) Did minority report get it wrong? Superiority of the mouse over 3D input devices in a 3D placement task. In: 12th IFIP TC 13 international conference on human–computer interaction: part II, Uppsala, Sweden. Springer, Berlin, pp 400–414

Bowman DA, Hodges LF (1997) An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. In: Symposium on interactive 3D graphics. ACM, pp 35–38

Bowman DA, Koller D, Hodges LF (1997) Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques. In: Virtual reality annual international symposium, 1997, IEEE 1997, 1–5 Mar 1997, pp 45–52, 215. doi:10.1109/vrais.1997.583043

Bowman DA, Johnson DB, Hodges LF (1999) Testbed evaluation of virtual environment interaction techniques. In: Paper presented at the proceedings of the ACM symposium on virtual reality software and technology, London, United Kingdom

Bowman DA, Kruijff E, LaViola J, Poupyrev I (2004) 3D user interfaces: theory and practice. Addison Wesley Longman Publishing Co., Inc., Boston

Callahan J, Hopkins D, Weiser M, Shneiderman B (1988) An empirical comparison of pie vs. linear menus. In: SIGCHI conference on human factors in computing systems, Washington, DC, United States. ACM, pp 95–100. doi:10.1145/57167.57182

Chen M, Mountford SJ, Sellen A (1988) A study in interactive 3-D rotation using 2-D control devices. SIGGRAPH Comput Graph 22(4):121–129. doi:10.1145/54852.378497

Coquillart S, Fuchs P, Grosjean J, Hachet M, Bechmann D, Stenberger L (2003) Les Techniques d'Interaction pour les Primitives Comportementales Virtuelles. In: Traité de la réalité virtuelle: volume 2, L'Interfaçage, l'Immersion et l'Interaction, vol 2, 3 edn. Les Presses de l'École des Mines de Paris, p 332

Darken RP, Sibert JL (1996) Wayfinding strategies and behaviors in large virtual worlds. In: Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems, Vancouver, British Columbia, Canada

Dewaele G, Devernay F, Horaud R (2004) Hand motion from 3D point trajectories and a smooth surface model. In: 8th European conference on computer vision, Prague, Czech Republic. Springer, pp 495–507

Dorner B (1994) Chasing the colour glove: visual hand tracking. Simon Fraser University, Burnaby

Elmezain M, Al-Hamadi A, Appenrodt J, Michaelis B (2008) A hidden Markov model-based continuous gesture recognition system for hand motion trajectory. In: 19th international conference on pattern recognition, Tampa, FL, USA. IEEE, pp 1–4

Fiorentino M, Radkowski R, Stritzke C, Uva A, Monno G (2013) Design review of CAD assemblies using bimanual natural interface. Int J Interact Des Manuf 7(4):249–260. doi:10.1007/s12008-012-0179-3

Fuchs P, Mathieu H (2003) Les Interfaces Spécifiques de la Localisation Corporelle - Introduction. In: Traité de la Réalité Virtuelle: volume 2, L'interfaçage, l'immersion et l'interaction, vol 2, 3 edn. Les Presses de l'École des Mines de Paris, pp 93–94

Fuchs P, Moreau G (2003) Traité de la Réalité Virtuelle, vol 2, 2 edn. Les Presses de l'Ecole des Mines de Paris

Geebelen G, Maesen S, Cuypers T, Bekaert P (2010) Real-time hand tracking with a colored glove. In: 3D Stereo media, Luik, Belgium

Gratzel C, Fong T, Grange S, Baur C (2004) A non-contact mouse for surgeon–computer interaction. Technol Health Care 12(3):245–257

Guiard Y (1987) Asymmetric division of labor in human skilled bimanual action: the kinematic chain as a model. J Mot Behav 9(4):486–517

Haik E, Barker T, Sapsford J, Trainis S (2002) Investigation into effective navigation in desktop virtual interfaces. In: Paper presented at the seventh international conference on 3D Web technology, Tempe, Arizona, USA

Hand C (1997) A survey of 3D interaction techniques. Comput Graph Forum 16(5):269–281

Hassanpour R, Shahbahrami A, Wong S (2008) Adaptive Gaussian mixture model for skin color segmentation. Int J Comput Inf Sci Eng 2(2):1–6

Hayward V, Astley OR (1996) Performance measures for haptic interfaces. In: 7th International symposium on robotics research. Springer, pp 195–207

Homma K, Takenaka E-I (1985) An image processing method for feature extraction of space-occupying lesions. J Nucl Med 26(1985):1472–1477

Hong D, Woo W (2006) A 3D vision-based ambient user interface. Int J Hum Comput Interact 20(3):271–284. doi:10.1207/s15327590ijhc2003_6

Hürst W, Wezel C (2013) Gesture-based interaction via finger tracking for mobile augmented reality. Multimed Tools Appl 62(1):233–258. doi:10.1007/s11042-011-0983-y

Jaehong L, Heon G, Hyungchan K, Jungmin K, Hyoungrae K, Hakil K (2013) Interactive manipulation of 3D objects using Kinect for visualization tools in education. In: Control, automation and systems (ICCAS), 2013 13th international conference on, 20–23 Oct 2013, pp 1220–1222. doi:10.1109/iccas.2013.6704175

Khan A, Mordatch I, Fitzmaurice G, Matejka J, Kurtenbach G (2008) ViewCube: a 3D orientation indicator and controller. In: Paper presented at the proceedings of the 2008 symposium on interactive 3D graphics and games, Redwood City, California

Kim Y, Leonard S, Shademan A, Krieger A, Kim PW (2014) Kinect technology for hand tracking control of surgical robots: technical and surgical skill comparison to current robotic masters. Surg Endosc 1–8. doi:10.1007/s00464-013-3383-8

Klein T, Guéniat F, Pastur L, Vernier F, Isenberg T (2012) A design study of direct-touch interaction for exploratory 3D scientific visualization. Comput Graph Forum 31(3pt3):1225–1234

Kolb A, Barth E, Koch R, Larsen R (2009) Time-of-flight sensors in computer graphics. In: Pauly M, Greiner G (eds) 30th annual conference of the European association for computer graphics, Munich, Germany, pp 119–134

Lange R (2000) 3D time-of-flight distance measurement with custom solid-state image sensors in CMOS/CCD-technology. University of Siegen, Siegen

Lange R, Seitz P (2000) Seeing distance—a fast time-of-flight 3D camera. Sens Rev 20(3):212–217

Laurel BK (1986) Interface as mimesis. User centered system design: new perspectives on human–computer interaction. Lawrence Erlbaum Associates, Hillsdale

LeapMotion (2012). https://www.leapmotion.com

Lempereur M (2008) Simulation du Mouvement d'Entrée dans un Véhicule Automobile. Université de Valenciennes et du Hainaut-Cambrésis

Levesque JC, Laurendeau D, Mokhtari M (2011) Bimanual gestural interface for virtual environments. In: Virtual reality conference

(VR), 2011 IEEE, 19–23 March 2011, pp 223–224. doi:10.1109/vr.2011.5759479
Lin J, Sun Q, Li G, He Y (2013) SnapBlocks: a snapping interface for assembling toy blocks with XBOX Kinect. Multimed Tools Appl 1–24. doi:10.1007/s11042-013-1690-7
Livingston MA, Sebastian J, Ai Z, Decker JW (2012) Performance measurements for the Microsoft Kinect skeleton. In: IEEE virtual reality, 2012. IEEE Computer Society, pp 119–120. doi:10.1109/vr.2012.6180911
Loup-Escande E, Burkhardt J-M, Richir S (2011) Anticiper et Evaluer l'Utilité dans la Conception Ergonomique des Technologies Emergentes: Une Revue. Le Travail Humain 76:27–55
Maes P-J, Amelynck D, Lesaffre M, Leman M, Arvind DK (2012) The "Conducting Master": an interactive, real-time gesture monitoring system based on spatiotemporal motion templates. Int J Hum Comput Interact. doi:10.1080/10447318.2012.720197
Mason AH, Bernardin BJ (2009) Vision for performance in virtual environments: the role of feedback timing. Int J Hum Comput Interact 25(8):785–805. doi:10.1080/10447310903025529
May S, Pervoelz K, Surmann H (2007) 3D cameras: 3D computer vision of wide scope. Int J Adv Rob Syst 4:181–202
McCrae J, Mordatch I, Glueck M, Khan A (2009) Multiscale 3D navigation. In: Paper presented at the proceedings of the 2009 symposium on interactive 3D graphics and games, Boston, Massachusetts
Mine M (1995) Virtual environment interaction techniques. UNC Chapel Hill CS Dept
Mistry P, Maes P, Chang L (2009) WUW—Wear Ur World: a wearable gestural interface. In: 27th international conference extended abstracts on human factors in computing systems, Boston, MA, USA. ACM, pp 4111–4116. doi:10.1145/1520340.1520626
Moerman C, Marchal D, Grisoni L (2012) Drag'n go: simple and fast navigation in virtual environment. In: 2012 IEEE Symposium on 3D user interfaces (3DUI), 4–5 March 2012, pp 15–18. doi:10.1109/3dui.2012.6184178
Mohr D, Zachmann G (2009) Continuous edge gradient-based template matching for articulated object. In: International conference on computer vision theory and applications, Lisbon, Portugal, pp 519–524
Mohr D, Zachmann G (2010a) FAST: Fast adaptive silhouette area based template matching. In: Labrosse F, Zwiggelaar R, Liu Y, Tiddeman B (eds) British machine vision conference. BMVA Press, pp 39.31–39.12
Mohr D, Zachmann G (2010b) Silhouette area based similarity measure for template matching in constant time. In: Proceedings of the 6th international conference on articulated motion and deformable objects, Mallorca, Spain. Springer
Movea (2009) MotionPod™ Technology. http://movea.com/healthcare/motion_pod/index.html. Accessed 26 May 2009
Nan X, Zhang Z, Zhang N, Guo F, He Y, Guan L (2013) vDesign: toward Image Segmentation and composition in CAVE using finger interaction. In: International conference on signal and information processing (ChinaSIP), Beijing, China, 6–10 July 2013. IEEE, pp 461–465
Nedel LP, Dal Sasso Freitas CM, Jacob LJ, Pimenta MS (2003) Testing the use of egocentric interactive techniques in immersive virtual environments. In: Paper presented at the 9 h IFIP TC13 international conference on human–computer interaction, Zurich, Switzerland
Norman DA (1988) The psychology of everyday things. Basic Books, New York
Ouhaddi H, Horain P (1998) Conception et Ajustement d'un Modèle 3D Articulé de la Main. In: 6èmes Journées de Travail du GT Réalité Virtuelle, Issy-les-Moulineaux, France, 13/03/1998 1998, pp 83–90
Pamplona VF, Fernandes LAF, Prauchner J, Nedel LP, Oliveira MM (2008) The image-based data glove. In: 10th symposium on virtual and augmented reality, João Pessoa, Brazil, pp 204–211
Pedersoli F, Benini S, Adami N, Leonardi R (2014) XKin: an open source framework for hand pose and gesture recognition using Kinect. Vis Comput 1–16. doi:10.1007/s00371-014-0921-x
Poor GM, Tomlinson BJ, Guinness D, Jaffee SD, Leventhal LM, Zimmerman G, Klopfer DS (2013) Tangible or gestural: comparing tangible vs. Kinect™ interactions with an object manipulation task. In: International conference on tangible, embedded and embodied interaction, Barcelona, Spain
Poupyrev I, Billinghurst M, Weghorst S, Ichikawa T (1996) The go-go interaction technique: non-linear mapping for direct manipulation in VR. In: Paper presented at the 9th annual ACM symposium on user interface software and technology, Seattle, Washington, USA
Poupyrev I, Ichikawa T, Weghorst S, Billinghurst M (1998) Egocentric object manipulation in virtual environments: empirical evaluation of interaction techniques. Comput Graph Forum 17(3):41–52. doi:10.1111/1467-8659.00252
Prisacariu VA, Reid I (2011) Robust 3D hand tracking for human computer interaction. In: IEEE international conference on automatic face & gesture recognition and workshops, pp 368–375
Raheja JL, Chaudhary A, Singal K (2011) Tracking of fingertips and centers of palm using KINECT. In: Computational intelligence, modelling and simulation (CIMSiM), 2011 third international conference on, 20–22 Sept 2011, pp 248–252. doi:10.1109/CIMSim.2011.51
Robinett W, Holloway R (1992) Implementation of flying, scaling and grabbing in virtual worlds. In: Paper presented at the proceedings of the 1992 symposium on interactive 3D graphics, Cambridge, Massachusetts, USA
Rodríguez N, Wikström R, Lilius J, Cuéllar M, Delgado Calvo Flores M (2013) Understanding movement and interaction: an ontology for Kinect-based 3D depth sensors. In: Urzaiz G, Ochoa S, Bravo J, Chen L, Oliveira J (eds) Ubiquitous computing and ambient intelligence. Context-awareness and context-driven interaction, vol 8276. Lecture Notes in Computer Science. Springer International Publishing, pp 254–261. doi:10.1007/978-3-319-03176-7_33
Schlattmann M, Klein R (2009) Efficient bimanual symmetric 3D manipulation for markerless hand-tracking. In: Paper presented at the virtual reality international conference, Laval, France
Schlattmann M, Na Nakorn T, Klein R (2009) 3D interaction techniques for 6 DOF markerless hand-tracking. In: International conference on computer graphics, visualization and computer vision, Plzen-Bory, Czech Republic
Shen Y, Ong SK, Nee AYC (2011) Vision-based hand interaction in augmented reality environment. Int J Hum Comput Interact 27(6):523–544. doi:10.1080/10447318.2011.555297
Soh J, Choi Y, Park Y, Yang HS (2013) User-friendly 3D object manipulation gesture using Kinect. In: Paper presented at the proceedings of the 12th ACM SIGGRAPH international conference on virtual-reality continuum and its applications in industry, Hong Kong, Hong Kong
Song P, Goh WB, Hutama W, Fu C-W, Liu X (2012) A handle bar metaphor for virtual object manipulation with mid-air interaction. In: Paper presented at the SIGCHI conference on human factors in computing systems, Austin, Texas, USA
Song J, Cho S, Baek S-Y, Lee K, Bang H (2014) GaFinC: gaze and finger control interface for 3D model manipulation in CAD application. Comput Aided Des 46:239–245
Stefano LD, Marchionni M, Mattoccia S (2004) A fast area-based stereo matching algorithm. Image Vis Comput 22(12):983–1005
Stenger B (2006) Template-based hand pose recognition using multiple cues. In: Computer vision, Hyderabad, India, 2006.

Lecture Notes in Computer Science. Springer, pp 551–560. doi:10.1007/11612704_55
Stenger B, Thayananthan A, Torr PHS, Cipolla R (2006) Model-based hand tracking using a hierarchical bayesian filter. IEEE Trans Pattern Anal Mach Intell 28(9):1372–1384. doi:10.1109/tpami.2006.189
Stoakley R, Conway MJ, Pausch R (1995) Virtual reality on a WIM: interactive worlds in miniature. In: Paper presented at the proceedings of the SIGCHI conference on human factors in computing systems, Denver, Colorado, USA
Theobalt C, Albrecht I, Haber J, Magnor M, Seidel H-P (2004) Pitching a baseball: tracking high-speed motion with multi-exposure images. ACM Trans Graph 23(3):540–547. doi:10.1145/1015706.1015758
Tokatli A (2005) 3D hand tracking in video sequences. Middle East Technical University, Çankaya
Tosas M (2006) Visual articulated hand tracking for interactive surfaces. University of Nottingham, Nottingham
Tosas M, Bai L (2007) Virtual touch screen: a vision-based interactive surface. In: 9th IASTED international conference on computer graphics and imaging, Innsbruck, Austria. ACTA Press, pp 81–86
Tricot A, Plégat-Soutjis F, Camps J-F, Amiel A, Lutz G, Morcillo A (2003) Utilité, utilisabilité, acceptabilité : interpréter les relations entre trois dimensions de l'évaluation des EIAH. In: Conférence EIAH 2003, Strabourg, France
Tubiana R, Kapandji A-I (1991) Affections Neurologiques. Traité de Chirurgie de la Main, vol 4. Masson, Paris, pp 56–58
Ueda E (2003) Hand pose estimation for vision-based human interface. IEEE Trans Industr Electron 50(4):676–684
Ughini CS, Blanco FR, Pinto FM, Freitas CM, Nedel LP (2006) EyeScope: a 3D interaction technique for accurate object selection in immersive environments. In: SBC symposium on virtual reality 2006, pp 77–88
Unseok L, Tanaka J (2012) Hand controller: image manipulation interface using fingertips and palm tracking with Kinect depth data. In: Asia Pacific conference on computer human interaction
Wang RY (2011) Practical color-based motion capture. Massachusetts Institute of Technology, Cambridge
Wang R, Paris S, Popovic J (2011) 6D hands: Markerless hand-tracking for computer aided design. In: Paper presented at the proceedings of the 24th annual ACM symposium on user interface software and technology, Santa Barbara, California, USA
Ware C, Osborne S (1990) Exploration and virtual camera control in virtual three dimensional environments. SIGGRAPH Comput Graph 24(2):175–183. doi:10.1145/91394.91442
Winer BJ (1971) Statistical principles in experimental design, 2nd edn. McGraw-Hill, New York City
Winkler S, Yu H, Zhou Z (2007) Tangible reality desktop for digital media management. In: Engineering reality of virtual reality, San Jose. SPIE
Yeo H-S, Lee B-G, Lim H (2013) Hand tracking and gesture recognition system for human–computer interaction using low-cost hardware. Multimed Tools Appl 1–29. doi:10.1007/s11042-013-1501-1
Zhang Z, McInerney T, Zhang N, Guan L (2014) A cave based 3D immersive interactive city with gesture interface. In: Paper presented at the 22nd WSCG international conference on computer graphics, visualization and computer vision, Plzen, Czech Republic, June 2–5, 2014
Zhou H, Hu H (2008) Human motion tracking for rehabilitation—a survey. Biomed Signal Process Control 3(1):1–18
Zhou R, Junsong Y, Jingjing M, Zhengyou Z (2013) Robust part-based hand gesture recognition using Kinect sensor. IEEE Trans Multimedia 15(5):1110–1120. doi:10.1109/tmm.2013.2246148