A Handle Bar Metaphor for Virtual Object Manipulation with Mid-Air Interaction

Peng Song **Wooi Boon Goh** **William Hutama** **Chi-Wing Fu** **Xiaopei Liu**
Nanyang Technological University, Singapore
{song0083,aswbgoh,hwilliam,cwfu,liuxp}@ntu.edu.sg

ABSTRACT
Commercial 3D scene acquisition systems such as the Microsoft *Kinect* sensor can reduce the cost barrier of realizing mid-air interaction. However, since it can only sense hand position but not hand orientation robustly, current mid-air interaction methods for 3D virtual object manipulation often require contextual and mode switching to perform translation, rotation, and scaling, thus preventing natural continuous gestural interactions. A novel *handle bar metaphor* is proposed as an effective visual control metaphor between the user's hand gestures and the corresponding virtual object manipulation operations. It mimics a familiar situation of handling objects that are skewered with a bimanual handle bar. The use of relative 3D motion of the two hands to design the mid-air interaction allows us to provide precise controllability despite the *Kinect* sensor's low image resolution. A comprehensive repertoire of 3D manipulation operations is proposed to manipulate single objects, perform fast constrained rotation, and pack/align multiple objects along a line. Three user studies were devised to demonstrate the efficacy and intuitiveness of the proposed interaction techniques on different virtual manipulation scenarios.

Author Keywords
3D manipulation; bimanual gestures; user interaction

ACM Classification Keywords
H.5.2 : User Interfaces - Input devices and strategies; I.3.6 : Methodology and Techniques - Interaction techniques

General Terms
Design

INTRODUCTION
In recent years, mid-air interaction supported by 3D spatial gestural inputs has received increasing attention from both the research community [8, 23, 17, 27, 4] and the gaming industry, as evidenced by the popular gaming devices such as Nintendo *Wii-mote* and Microsoft *Kinect*, which allow us to perform natural physical interactions in our own physical space while moving freely in front of a large display.

CHI'12, May 5–10, 2012, Austin, Texas, USA.

There are basically two approaches to accommodate mid-air interactions in such a visual interactive setting. The first employs a handheld controller device, such as the Nintendo *Wii-mote*. User inputs via button clicks and accelerometer-based motion sensing are integrated to form high-level gestures to support the interaction. The second is a controller-free approach, where users can manipulate the graphical contents on the display with their bare hands. Temporal information to support mid-air interaction is obtained by using an image and/or depth sensor (e.g., *Kinect*) to continuously sense and analyze the user's body posture and hand gestures via real-time image processing techniques.

This paper studies mid-air interaction designs to support object manipulation in a 3D virtual environment in a controller-free setting. This approach is particularly useful for distant viewing and interaction in front of large displays since users can directly perform spatial gestures in their own physical space. This physical space alone can be used to produce natural 3D manipulation inputs without cumbersome handheld peripherals. With the wide availability of the Microsoft *Kinect* sensor [26], the cost barrier of realizing such mid-air interaction system has been significantly reduced. However, due to the limitation of the *Kinect* sensor, which can robustly sense hand position but not hand orientation, current interaction methods often require mode switching to move between different operations such as rotation, translation and scaling. As a result, it is difficult for users to recall and execute these operations. This paper aims to address such shortcomings within a controller-free environment that supports natural and intuitive mid-air interactive gestures.

Figure 1. Manipulating a turkey with a bimanual handle bar.

At the heart of this inquiry is the question of what suitable metaphors one can use to map the 3D gestural actions of a user to the manipulation operations on objects in a 3D virtual environment. The metaphor we proposed for visual manipulation tasks is a bimanual handle bar shown in Figure 1. We call this the *handle bar* metaphor. Both hands from the users are employed to manipulate the virtual objects in a natural manner. After the related work section, we give an overview of the interaction system, and then describe the handle-bar-

based interaction designs for typical object manipulation operations in a 3D virtual space and highlight their advantages. User evaluations were carried out on various visual manipulation tasks that involve translation and rotation, constrained rotation, and multiple object alignment. Results show that all users can quickly improve their competency in performing the required tasks using our interaction design with only a short period of practice.

RELATED WORK

This section surveys various interaction paradigms to manipulate 3D objects in virtual environments and relevant mid-air interaction applications with the *Kinect* sensor.

Interaction with 3D Virtual Environments

There are a wide range of methods to interact with 3D contents in virtual space [9]. Since this work focuses on interactions with freehand gestures, we review mainly two more relevant areas: virtual reality and freehand interfaces.

Virtual Reality Interfaces. This approach immerses users in a virtual space for them to perform interaction via various sensors and input devices. Duval et al. [13] proposed a 3D interaction technique called "SkeweR," which enables two users to move the same virtual object collaboratively. John et al. [21] employed hand and head reconstruction as well as tracking for 3D interaction in a desk-based computer environment. More recently, Ang et al. [3] proposed to enable multi-point haptic grasp in a virtual environment by using a gripper attachment while Jacobs and Froehlich [20] developed a soft hand model to achieve robust finger-based manipulation of virtual objects.

Among the virtual reality interfaces, some employ data gloves for gestural mid-air interactions. Cutler et al. [12] built a virtual reality system that allows users to naturally manipulate virtual 3D models with both hands on a tabletop stereo display. In particular, they proposed a grab-and-carry tool for a user to hold an object with two hands, as well as to "carry" it and turn it around. Zigelbaum et al. [34] presented g-stalt, a gestural interface for users to navigate and manipulate a 3D graphical environment filled with video media using various hand gestures. Lévesque et al. [24] proposed a 3D bimanual gestural interface using data gloves for 3D environment interaction; the left hand is employed to perform gestures for selecting interaction modes while the right hand is for the interaction itself, e.g., rotating or scaling the desired object. Though VR interfaces provide highly immersive perception and interactive controls to users, they typically require users to wear instrumented gloves for gestural input, which could be uncomfortable and restrict the freedom of movement.

Freehand Interfaces. Freehand interfaces employ tracking systems to recognize mid-air hand or arm gestures as user input. Sato et al. [30] estimated 3D hand poses and recognized hand shape patterns in real-time using multiple cameras. Grossman et al. [15] developed interesting gestural interactions with multiple fingers over a spherical volumetric display. Luo and Kenyon [25] employed scalable computing methods for vision-based gesture interaction in a large display setting. Hilliges et al. [18] enabled intuitive manipulation of 3D digital contents by leveraging the space above the surface of a regular interactive tabletop display. Benko and Wilson [7] proposed to interact with a large curved display by combining speech commands with freehand pinch gestures to provide immersive and interactive experience to multiple users. More recently, Nancel et al. [27] proposed a set of mid-air gestures to support pan-and-zoom interaction with graphical contents shown on a wall-sized display.

To manipulate a 3D virtual object with a single hand, one typical metaphor is to grip and manipulate it with the thumb and forefinger, i.e., a pinch gesture. Segen and Kumar [31] described the GestureVR system that used this metaphor to continuously manipulate 6DOF of an virtual object; the object can be translated by moving the hand and oriented by rotating the wrist. O'Hagan et al. [28] later extended this metaphor by allowing users to resize the object by moving the thumb and forefinger apart or towards each other. Though this metaphor is very natural and intuitive for common users, it requires fine and robust detection of dynamic fingers poses, which is not achievable with the poor image resolution of low-cost depth sensing devices such as the *Kinect* sensor, or when the user stands too far away from the sensor as in the case of large display setting.

Closer to our approach are freehand interactive systems that employ two-handed gestures, i.e., bimanual interaction [16]. Hinckley et al. [19] discussed two-handed user interface design issues for 3D manipulation and highlighted the scientific measurement of human behavioral principles in the interface design while Brandl et al. [10] later compared bimanual interfaces with different combinations of pen and touch on a horizontal display. Benko and Wilson [6] enabled users to visualize and manipulate 3D virtual objects using bimanual gestures on an interactive surface and above it (in mid-air). However, their approach is unable to support simultaneous multiple object manipulation operations. Yoo et al. [33] combined the gaze and the hand gestures for manipulating 3D digital contents shown on a large-scale display; 3D bimanual gestures are used for virtually manipulating a collection of image elements. Hackenberg et al. [17] presented a freehand 3D multi-touch interface for 3D object manipulation using a time-of-flight camera; a 3D object can be translated by moving one hand in 3D space while the object can be rotated and scaled simultaneously using a two-touch-point metaphor. Compared to these approaches, our use of a handle bar that can be positioned outside the selected object allows us to flexibly perform non-object centric manipulation. Very recently, Wang et al. [32] developed a bimanual interaction system for assembling CAD components by using two webcams to track 6 DOF of each hand, where a sheet-of-paper metaphor is proposed for performing the rotation. Unlike [32], our handle bar metaphor design seeks to provide precision control for all R-T manipulations in a unified bimanual manner. In addition, it does not combine mid-air and keyboard interaction since it is designed for use in a "in-front-of large display" setting with a low-resolution tracking system such as the Kinect sensor.

Interaction with the Kinect sensor

Kinect [26] is a controller-free real-time depth sensing device, primarily designed for supporting gaming with the Mi-

crosoft Xbox360 system. Since its launch, it had sold at an average volume of around 133 thousand units per day in its first sixty days. Due to its low-cost and wide availability, it has not only gained popularity for gaming, but also employed in numerous research projects in various disciplines. In particular, this recent innovation spawned many interesting mid-air interaction applications, which have made their rapid debut on the Internet. For example, the manipulation of 2D and 3D objects [11, 22], tracking of human motions, gesture control for robot systems, multi-touch-like interface for controlling GUI functions like those seen in *Minorty report*, see [2], and [1]. In this work, we explored the use of this low-cost device for object manipulation. Our proposed handle bar design can support efficient and effective bimanual manipulation of 3D objects while accommodating the limitations posed by the *Kinect* sensor.

SYSTEM SETUP

Our system setup consists of an Alienware Aurora ALX desktop computer with QuadCore CPU 3.20GHz and 9GB memory, running Linux Ubuntu 10.10 (Maverick) with an NVIDIA 1.5GB GeForce GTX480 graphics board, a *Kinect* sensor, supporting an image resolution of 320×240 at 30 frames per second with both color and depth, and an LCD display of physical size 32 inches. The *Kinect* sensor is placed below the large display and the user stands at a distance of around 2 meters from the display during the interaction (see Figure 2).

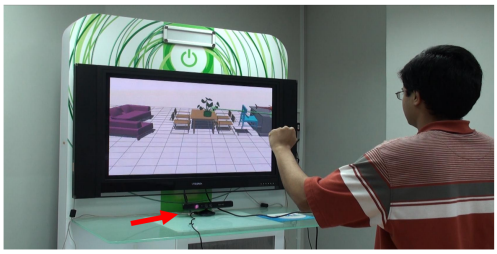


Figure 2. Our system setup with the *Kinect* sensor (red arrow).

Software. We use PrimeSense's OpenNI [29] open source drivers and the NITE middleware to interface with the *Kinect* sensor; the depth generator in the OpenNI framework is first employed to obtain the depth image data from *Kinect*. Then, we use the skeleton tracker in NITE to compute the user's joint positions from the depth image so that we can determine the 3D location of the user's hands. At the same time, we use the perception point cloud library (PCL) from the Robot Operating System (ROS) framework [14] to generate point clouds from the depth image. Lastly, based on the hand locations obtained from the 3D skeleton, we segment a point cloud set associated with each of user's hands. Our experience suggests that the use of the 3D skeleton as a guide produces more accurate and robust segmentation.

Hand Gesture Recognition. Our system is able to recognize three basic single-handed gestures, namely POINT, OPEN, and CLOSE (see Figure 3). To differentiate among them, the extracted point cloud data of each hand is first low-pass filtered (over 30 frames) to remove unintentional hand shaking. These segmented clusters of unity-weighted points [14] permit two 3D centroid locations to be computed. The spatial distribution of the points in the point cloud (after offset by the centroid) is then computed and pattern-matched with

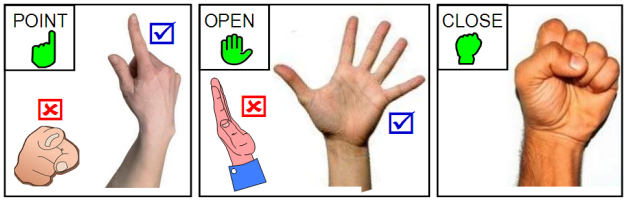


Figure 3. The different recognized hand gestures as seen by the 3D scene acquisition sensor and their respective visual icons used on the large display for visual feedback. The (a) POINT finger and (b) OPEN palm gestures are less stable as their shapes will change based on the orientation of the hand, (c) unlike the CLOSE fist gesture.

the point distributions of the three known gesture classes to determine which hand gesture is currently active. If a hand is located below the center of the user's body, a DOWN gesture is assigned to the hand. This allows the system to distinguish between one and two-handed interactions. In addition, the two centroid points from each of the two hands (computed at a rate of 30 frames-per-second) give the instantaneous length and 3D orientation of the handle bar.

THE HANDLE BAR METAPHOR

Consider the task of manipulating a 3D virtual object on a wall display using only our two bare hands. What would be the most effective and intuitive way to do this? A survey of existing literature revealed a dearth of mid-air interactive designs to perform such a task, especially in environments where multiple objects can be independently manipulated.

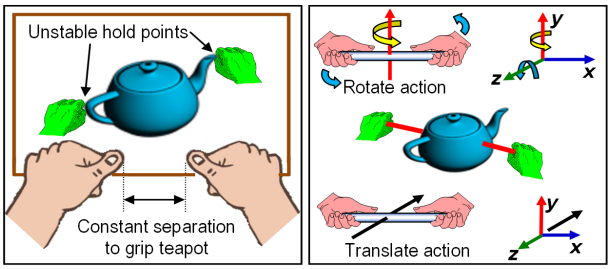


Figure 4. (a) The metaphor of *two remote gripping-hands* projected into the 3D virtual space, which is shown gripping a teapot. (b) The metaphor of a *handle bar* extended from two clasp hands, which is used to pierce through the teapot for rotation and translation manipulations.

One possible approach is to project our physical hands into the virtual space using two iconic hands that represent the hand positions and gestures (see Figure 4 (a)). Using the iconic visual feedback, the user can move one's hands to grip the virtual object and then rotate-translate (R-T) it with further coordinated hand movements. This *two remote gripping-hands* metaphor has a direct representation in the virtual space and provides a good semantic mapping between the physical and virtual manipulation. However, without haptic feedback, it demands substantial physical dexterity to maintain the gripping separation whilst performing the basic R-T manipulations. Moreover, the hand icons can be easily occluded by the virtual object during the rotation. Direct grip-based metaphors can also be problematic because the virtual object may not have stable flat contact surfaces for gripping.

To overcome these limitations, a novel *handle bar* metaphor is proposed. In this metaphor, we pierce a virtual handle

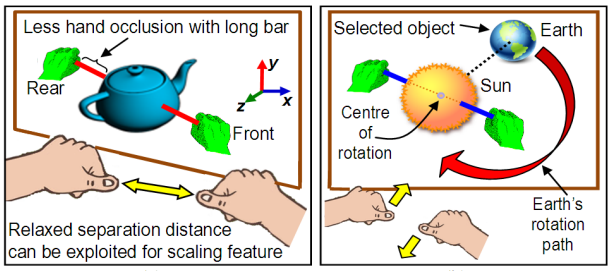


Figure 5. Features of the *handle bar* metaphor. (a) Scaling operations done by varying distance between the two hands. (b) Rotation of a selected object (Earth) performed about the center of the handle bar placed inside another object (Sun).

bar through the selected 3D object. With the object now attached to the handle bar, manipulation of the object is done by performing R-T manipulations on the handle bar instead (see Figure 4 (b)). Unlike the *two remote gripping-hands* approach, the handle bar icon (the red line in Figure 5 (a)) provides helpful visual feedback to the user, continuously presenting the relative orientation of the two hands in 3D space during our interactive manipulation. A summary of the advantages of the proposed handle bar metaphor as suitable interaction paradigm for mid-air interaction is as follows:

- *Physical familiarity* - Bimanual motion gestures required to manipulate the handle bar are intuitive for most users since holding and manipulating an elongated bar with our two hands is a familiar undertaking in common activities such as cycling and lawn mowing.
- *Rich variety of 3D manipulation operations* - The handle bar interaction design offers seamless 7DOF manipulation (3 translations, 3 rotations, and 1 scaling), allowing users to transit smoothly between operations such as translation, rotation, and even scaling (see Figure 5 (a)), without changing gestures or operational modes. Interaction design for fast and precise constrained rotation can also be realized with a perpendicular extension to the virtual handle bar metaphor. Speedy multi-object manipulation can also be supported by piercing the handle bar through more than one virtual object. These pierced objects can be made to align or slide along the handle bar by using simple variations to the standard bimanual gestures.
- *Supporting both object and non-object centered manipulations* - By allowing the user to manipulate the position of the handle bar to any location relative to the selected 3D object, manipulation of virtual objects need not be object-centric. Figure 5 (b) shows an example that has a selected object (Earth) rotated about another (Sun).
- *Good semantic mapping* - Unlike other two-handed interaction methods [27] that combine different hand gestures to realize a larger subset of operations, the handle bar metaphor inspires bimanual gestures that have good semantic mapping to the physical world: the handle bar is "grabbed" for manipulation by clutching two fists and is "released" with open palms; pointing-finger gesture (finger prodding analogy) is used to change the position and orientation of the virtual handle bar (see Figure 6 (a)).
- *Accommodating sensor limitations* - With the limited resolution of the *Kinect* 3D scene acquisition sensor, stable and accurate 3D pose information can only be reliably specified with end points that are separated by some distance. The handle bar metaphor circumvents this deficiency by designing object manipulation controls that are based on the manipulation of an elongated bar that is specified by the two separated hands of the user.

INTERACTION DESIGN

This work is concerned with enabling a single user to interactively manipulate single or multiple 3D objects in a virtual environment. Users execute different visual manipulation operations by moving one or two hands freely within the physical space defined by their frontal bimanual arm-reach envelope. This section discusses the handle-bar-based interaction designs to perform the three basic categories of manipulation operations summarized in Figure 6. One manipulates the handle bar (see Figure 6 (a)). Another involves the manipulation of both single and multiple virtual objects (see Figure 6 (b,c)), and the last one manipulates the view of a virtual camera in the 3D environment (see Figure 6 (d)).

Hand Gestures Design

Our system can recognize three basic hand gestures, namely POINT, OPEN, and CLOSE (see Figure 3). As highlighted in Figure 6, the interaction design employs a consistent interpretation of these hand gestures. The POINT and CLOSE gestures are always associated with the handle bar and virtual object, respectively. Homogenous bimanual gestures will perform basic rotation-translation-scaling (RTS) manipulation of the handle bar or object, depending on whether

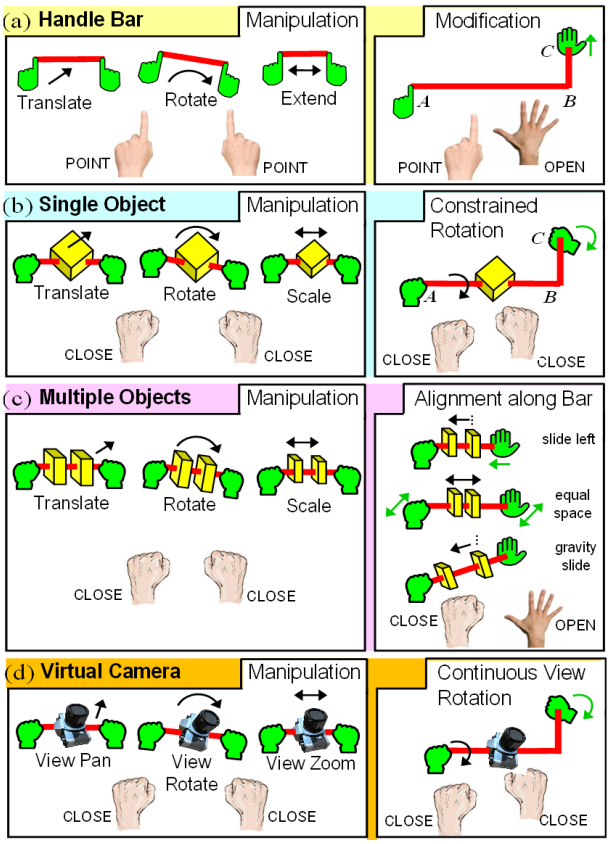


Figure 6. The various operations designed for the manipulation of (a) the virtual handle bar, (b) a single object, (c) multiple objects, and (d) the virtual camera and their associated bimanual hand gestures.

POINT or CLOSE gestures are used. Combining POINT and OPEN gestures allows the handle bar to be modified for constrained rotation (see Figures 6 (a)). A combination of CLOSE and OPEN gestures allows multi-object alignment along the handle bar (see Figure 6 (c)).

As shown in Figure 3 (a,b), both the POINT and OPEN hand gestures are sensitive to viewpoint changes, which often make their automatic recognition and classification less robust than the CLOSE gesture. Hence, they are assigned to interactions that are gesturally less complicated and used less frequently, e.g., browsing and handle bar manipulation. Since the centroid computation of the CLOSE fist gesture is orientation-invariant and thus more stable, it is used in the object manipulation interactions that often require the user to perform bimanual motion gestures with high degree-of-freedom. This assignment also fits well into the semantic mapping of how a physical handle bar can be manipulated. Figure 7 shows the state transitions and the expected hand gestures at each state when a user manipulates a single object. Details of various states are described next.

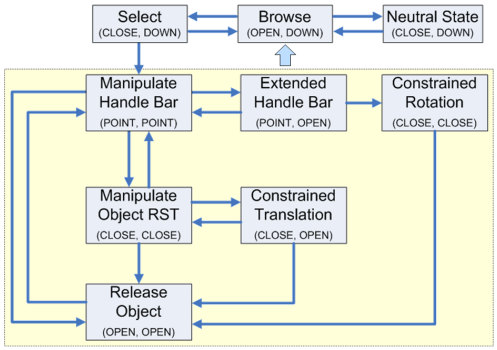


Figure 7. The state transition diagram for single object manipulation and the associated bimanual gestures for each state.

Neutral State

This is the initial state when the system starts. In this state, no object or camera is selected, and as such, no manipulation can take place. The provision of a Neutral state is important as it helps overcome the immersion syndrome [5], where every hand gesture is captured and constantly interpreted by the system. This can lead to undesirable operations due to misinterpretation of the user's unintended hand gestures. When interaction is no longer desired, we can re-enter the Neutral state by selecting an empty screen region.

Browse and Select

The users leave the Neutral state and enter the Browse mode by keeping one hand on their side (unimanual gesture) and waving the other raised OPEN palm. A small open hand visual icon in the virtual space moves within a 2D plane in tandem with the movement of the raised OPEN palm (see Figure 8 (a)). When the open hand icon overlaps with a 3D object or the virtual camera icon, the user can perform a CLOSE hand gesture to select the item (see Figure 8 (b)). Upon selection, a virtual handle bar will protrude out of the object in the default orientation, namely through the object's centroid and along the x-axis. This virtual handle bar icon indicates that the system is no longer in Neutral state and is currently in the Selected state (see Figure 8 (c)). Multiple virtual objects can be selected by repeating this selection operation. An active item will be deselected by performing a CLOSE hand gesture when the open hand icon overlaps with the selected object. All the active items will be deselected if the user performs a CLOSE hand gesture when the open hand icon overlaps with an empty space. In the Selected state, we can then proceed to other interaction modes such as the mode to manipulate the virtual handle bar.

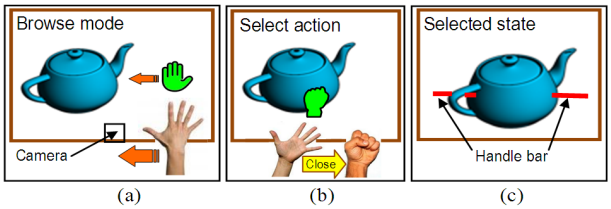


Figure 8. (a) Browsing the 3D virtual environment with a single OPEN palm in *Browse* mode. Selectable objects also include the virtual camera icon located at the bottom of the screen. (b) The object under the hand icon is selected with a Select action (CLOSE hand gesture). (c) In the Selected state, a handle bar protrudes out of the selected object.

Basic RTS Operations of a Single Object

The handle bar metaphor provides 7DOF manipulation (3D translation, 3D rotation, and 1D scaling) of virtual object and supports continuous transitions between operations. Such manipulation involves the appropriate placement of the virtual handle bar and the subsequent manipulation of the selected object about the center of the positioned bar. The modes associated with this process are described here. Note that we use different handle bar colors as a visual feedback to indicate which mode is currently active.

Manipulate Handle Bar Mode

Employing the bimanual POINT gesture in Figure 6 (a), users can manipulate the virtual handle bar by changing the relative position and orientation of the invisible line that joins their two hands in the physical space (see Figure 9 (a)). The end points of the handle bar are determined by the centroid of the 3D point clouds associated with the two POINT hand gestures. These were observed to be more stable end points than the more appropriate pointing finger tips, which result in handle bar jittering when switching between handle bar and object manipulation modes.

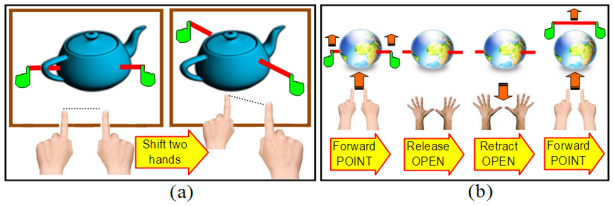


Figure 9. (a) Manipulate the 3D position and orientation of a handle bar using two pointed fingers. (b) Repeated translate (POINT gesture) and release (OPEN gesture) can be used to position handle bar at a distant away from the selected object.

The handle bar position is not restricted to the confines of the 3D object. Large translation of the handle bar can be achieved by repeatedly releasing the bar with a bimanual OPEN gesture, retracting the open palm, and then translating the bimanual POINT gesture again in the same direction (see Figure 9 (b)). In other words, the 3D gestural workspace need not have an absolute mapping to the 3D virtual world.

The user's physical translational motion can move the handle bar relative to its current 3D virtual world position. This convention is applied generically to the R-T interactions of the handle bar, selected 3D object, and the virtual camera.

The midpoint of the handle bar is the center for rotating the selected virtual object. During the handle bar manipulation, the selected object is not affected so that we can change the rotation center by translating the handle bar. Once the handle bar has been manoeuvred into the desired position, the user can manipulate the selected object by the next mode.

Manipulate Object Mode

The manipulate object mode is a bimanual interaction mode that employs two CLOSE fist gestures (see Figure 6 (b) left). We can apply three basic manipulation operations to a selected object: rotation, translation, and scaling (RTS).

Object Translation. The selected object can be translated in the x, y, and z directions by simply translating the bimanual CLOSE fist gesture in the corresponding directions (see Figure 10 (a)). The translation of the object is based on the movement vector of the virtual handle bar mid-point, as defined by the centroids of the two CLOSE fists.

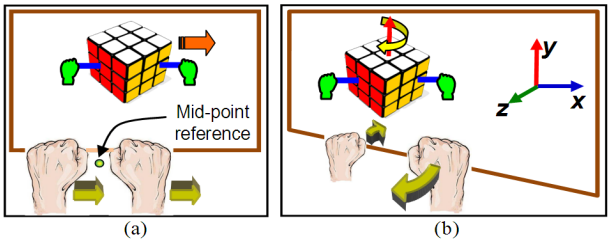


Figure 10. Bimanual CLOSE fist gestures for (a) translating and (b) rotating an object in the x and y-axis, respectively.

Object Rotation. The rotation of the selected object about a specific axis is based on the relative angular displacement of the virtual handle bar along that corresponding axis (see Figure 10 (b)). No absolute angular mapping is needed since the virtual handle bar can be released in a similar fashion as described in Figure 9 (b) (i.e., with OPEN palm gestures). Once released, the user may re-initiate a bimanual CLOSE fist gesture at a new position and perform a further rotation. This manner of executing a rotation allows the user to make large angular changes to the 3D virtual object about the y-axis without getting into an undesirable situation where the front hand occludes the back hand (see Figure 11 (a)), which may result in an indeterminable 3D pose of the virtual handle bar. Rotations about the x-axis cannot be directly manipulated using the handle bar since the wrist-based rotation of the two CLOSE fists does not change the position of their centroids and thus gives no angular rotation cues. However, rotation about the x-axis can still be obtained in an incremental fashion by executing appropriate concurrent bimanual uni-directional rotation about the y and z axes simultaneously (a "pedaling" motion) (see Figure 11 (b)), which is not immediately intuitive for uninitiated users. In this case, the constrained rotation provision (see Figure 6 (b) right) may be a better option as it provides faster and more precise manipulation albeit requiring a mode switch step.

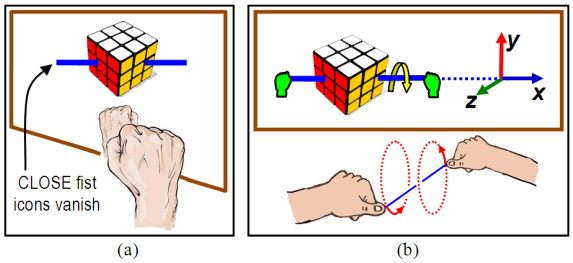


Figure 11. (a) Invalid gesture due to occlusion. (b) Incremental x-axis rotation using continuous rotation in the y and z axes.

Object Scaling. The object scaling operation is done by moving the two CLOSE fists towards each other (scale down) or away from each other (scale up), along the invisible line that connects the centroids of the two hands in physical space (see Figure 12). The scaling factor S is given by

$$S = \frac{\Delta L^2}{\Delta t}\ , \tag{1}$$

where ΔL is the change in distance between the two centroids in the sampling time Δt determined by the *Kinect* sensor's frame rate. In other words, the amount of scaling can be controlled by both the hand movement distance and speed. A vigorous gesture gives a larger scaling factor.

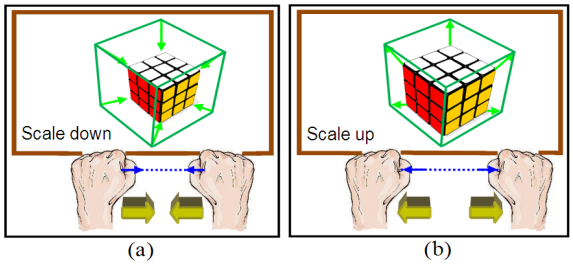


Figure 12. Object scaling gestures. Scaling (a) down and (b) up.

Constrained Operations of a Single Object

The ability to perform continuous translation and rotation can sometimes make it difficult to execute fast and precise rotation of an object about a specific straight line. In such circumstance, a less flexible constrained rotation operation could be more preferable. In our handle bar interaction design, constrained operations can be initiated with a combination of non-homogenous bimanual gestures.

Constrained Rotation. From a two POINT finger handle bar manipulation gesture (see Figure 6 (a) left), the user can change one hand to an OPEN palm gesture and move the OPEN palm away from the handle bar axis to create a "cranking bar" with a perpendicular extension (see Figure 6 (a) right). In detail, the horizontal line AB is defined by the standard handle bar when the palm OPENs. After the user moves the right OPEN palm to define the vertical line BC (see Figure 6 (a) right), one can CLOSE both fists to enter the constrained rotation state. To execute the constrained rotation, the user holds the left CLOSE fist steady and performs a "cranking" action with the right CLOSE fist about the virtual line AB. The angular velocity can be controlled by the length of virtual line BC, which is drawn continuously on the display as a helpful visual feedback. A shorter BC extension gives faster rotation but less precise angular

positioning while a longer BC extension gives better control of angular position at the expense of faster rotation.

Constrained Translation. Albeit less useful, constrained translation of a single object along the handle bar can be performed with a non-homogenous combination of a CLOSE fist and an OPEN palm. Sliding the OPEN palm towards the CLOSE fist moves the single object on the handle bar towards the CLOSE fist end. This idea is more useful when applied to the manipulation of multiple objects.

Manipulation and Alignment of Multiple Objects

A group of objects can be manipulated together and/or aligned along the virtual handle bar once they are selected (using repeated selection actions) and are all pierced by a handle bar.

Manipulation of Multiple Objects. First, standard RTS operations can be performed on all these selected objects in the same manner as with a single object (see Figure 6 (c) left). All objects on the handle bar can translate and scale as an aggregation. Rotation of all these objects will be centered about the mid-point of the handle bar.

Alignment of Multiple Objects. Three basic alignment operations for aligning multiple objects on a handle bar are provided (see Figure 6 (c) right):

- The first allows the user to "pack" objects by interactively sliding the selected objects towards each other using a gesture that moves the OPEN palm towards the CLOSE fist. Objects stop sliding when boundary collision is detected. Multiple objects can also be made to slide towards the CLOSE fist by "tilting" the virtual handle bar as shown in Figure 13. The larger the tilt angle is, the faster the objects will slide. This manner of packing multiple objects has a very close semantic mapping to the physical nature of object behavior along a handle bar (under gravity) and may be preferred by some users.

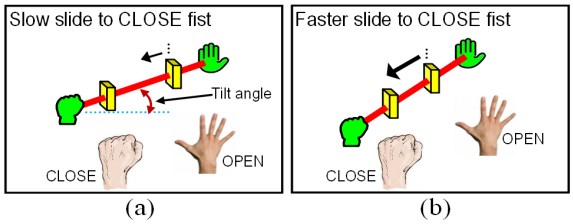


Figure 13. Gravity-simulated multiple object alignment. (a) Slow drop with gentle tilt. (b) Fast drop with steeper tilt.

- The second category of alignment operation is for evenly-distributing the objects along the virtual handle bar. The user can shake both CLOSE and OPEN hands and objects on the handle bar will be distributed at equi-distance along the handle bar. This operation is useful for "unpacking" objects that are too close to each other. The separation distance can be controlled by the user by adjusting the length of the handle bar before the shaking.

EVALUATION

Three user studies were conducted to evaluate the various handle bar interaction designs. The complete gesture set was active in all the three user studies except object scaling, which was disabled as it was not required. Twelve participants (8 males and 4 females) aged between 21 to 28 years were enrolled. None of them has performed mid-air visual manipulation before but ten of them have played games with the *Kinect* sensor or *Nintendo Wii-mote* gaming system. The physical setup used in the study is shown in Figure 2. Before the start of each session, the required task was first explained to the user, and then, an expert user showed a demonstration of how the task could be done.

User Study: R-T Manipulation

One of the strengths of the interaction design using the proposed handle bar metaphor is the ability to execute continuous transitions between RTS manipulation operations without the need to switch modes. We wanted to evaluate if naive users were able to perform simple R-T manipulations without any training and whether subsequent repeated attempts can quickly improve their performance. Figure 14 (a) shows the task of rotating and translating a randomly-oriented cube to its desired position as indicated by the wire-frame outline. Before starting the task, a brief demonstration was given to show the user the required hand gestures to select an object, position a handle bar, and perform the R-T manipulation required to put the cube into its destination. At each attempt, the user was given 2 minutes to put as many cubes as possible into the wire-frame envelope subject to a reasonable precision indicated by a wire-frame color change.

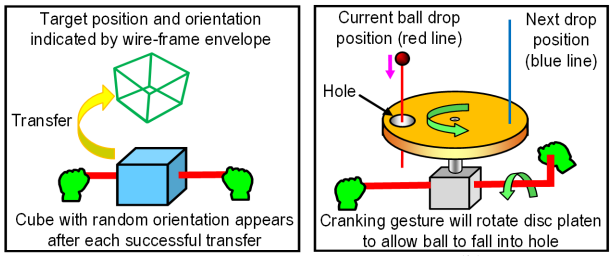


Figure 14. (a) User study: R-T Manipulation and (b) User study: constrained rotation.

Results in Table 1 show that the handle-bar-based R-T manipulation can be quickly learnt by just having on-the-task practice. On average, the 12 participants were able to double the number of cube placements within 6 attempts. However, the variance in user performance is high, indicating that some users are better in performing this type of interaction than others. The best performer managed 10 cubes in attempt #1 and improved to 15 cubes by attempt #6, compared to the worst performer who managed only 1 cube in attempt #1 but did improve to 6 cubes by attempt #6.

Attempt no.	#1	#2	#3	#4	#5	#6
Average times	4.6	5.9	7.0	8.1	8.3	9.3
Variance	7.2	8.7	13.8	12.6	9.0	10.8

Table 1. Results of the R-T manipulation user study.

User Study: Constrained Rotation

In this study, users were asked to perform constrained rotation about the x-axis (see Figure 6 (b) right). The task shown in Figure 14 (b) cannot be easily done with standard

rotation due to the possibility of inter-hand occlusion. The task evaluates the speed the user can rotate the disc clockwise or anti-clockwise to reach the desired angular position to "catch" the falling ball. To achieve this, the ball dropping speed increases linearly in each successive drop. This task also evaluates the angular precision the user can maintain in order to ensure the ball "drop" into the hole on the disc. For this, the hole is made small and a red vertical line provides the user with the visual cue required to align the "catch." The task is to "catch" as many falling balls as possible into the hole on the rotating disc within 60 seconds. Like before, a demonstration on how the task is done was first given to each user and attempt #1 was done without any practice. The subsequent two attempts were performed after giving the participants 2 minutes practice time before each attempt.

Attempt no.		#1 (No practice)	#2	#3
Number of balls caught	Average	11.7	17.6	17.3
	Variance	18.7	25.1	12.7
Ball count when 1st miss occurred	Average	3.7	8.8	9.3
	Variance	8.1	39.3	17.2

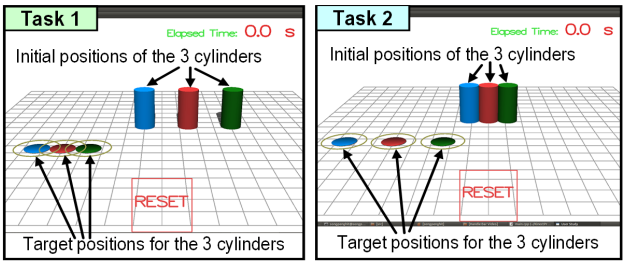
Table 2. Results of the constrained rotation user study.

Table 2 shows the average number of balls caught by the 12 participants; after a short period of practice time. The performance can increase from about 11 (attempt #1) to about 17 balls (attempts #2 and #3). The handle bar interaction design for rotating a virtual object about a fixed axis is able to provide fast angular speed, yet still offering good angular position controls since the speed of the 17th dropping ball is significantly faster than the 11th ball. This conclusion is further supported by the fact that on average, the first error (missed ball catch) made by the users were delayed from about the 3rd ball (attempt #1) to the 9th ball (attempts #2 and #3) after a short period of practice, again indicating the angular controllability and precision of the "cranking" bimanual CLOSE fist gesture despite the small room for angular error in catching the ball.

It is interesting to note the performance variability among the 12 participants after practice, as evident in the large variance increase in attempt #2 for both the number of balls caught and first ball missed. Like the first user study, this suggests that some users found executing the correct bimanual mid-air gesture requires more practice than just 2 minutes. The variance was observed to drop significantly after they were given a little more time (i.e., 2 more minutes) to practice the constrained rotation gesture.

User Study: Multiple Object Manipulation and Alignment

The final user study compares the time performance of manipulating three in-line objects such that we have to move them to some target positions. In Task 1, the three objects are initially positioned at a distance apart and needs to be brought close to one another at the destination. Task 2 does the reverse (see Figure 15). The users were asked to perform these manipulation tasks using the single object manipulation technique as well as the multiple object manipulation and alignment techniques (see Figure 6 (c)).



(a) (b)

Figure 15. This study measures the time taken to (a) bring objects together (Task 1) and (b) move objects apart (Task 2) using single and multiple object manipulation techniques.

Attempt no.		#1 (No practice)	#2	#3
Task 1 Avg time (sec)	One-by-one	53.7	37.0	34.5
	Multi-object	16.6	14.4	10.5
Task 2 Avg time (sec)	One-by-one	35.3	30.0	24.9
	Multi-object	18.2	13.3	13.5

Table 3. Results of the multiple object interaction user study.

Results in Table 3 show that for both Tasks 1 and 2, it was at least 2 to 3 times faster when using the multi-object manipulation and alignment techniques to do the required task than placing objects one at a time. From the absolute average time taken and rate of improvement with each subsequent attempt, it is clear that the "pack" multi-objects alignment procedure of Task 1 is easier to execute than the "equi-distribute" multi-object alignment procedure of Task 2.

Discussion and Limitations

A set of questionnaires were given to each user after each of the three user studies to gauge the subjective aspects of their experiences in the handle bar interaction. Table 4 presents the mean response values of the 12 participants and the associated standard deviation bar for each survey question.

The bimanual R-T manipulation hand gesture designed using the handle bar metaphor was found by most users to be generally intuitive to use, ease to remember, and providing good controllability. Consistently high ratings were received from the users for the ability to perform rotation and translation in one continuous motion. The subjective evaluation of the constrained rotation interaction design fared a little worst, with mean ratings at values just below 4.0. The user variability was far higher though, suggesting that performing fast and precise angular rotations with a cranking action is not universally straightforward for everyone. The very high mean rating for question #8 suggests that most users find the ability to rotate, translate, and align multiple objects at the same time to be very useful and preferred when manipulating several objects with similar trajectory and orientation.

Some limitations of the handle bar design were observed from the user studies: 1) Some users complained about arm fatigue after 20-30 minutes of continuous usage, which seems to be a universal drawback for all mid-air interaction designs that require precise control of the hands but provide no additional physical support for the extended arms; 2) User mem-

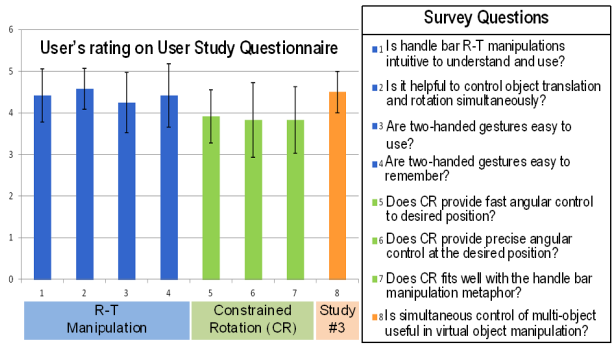


Table 4. User ratings on a Likert scale from 1 (strongly disagree) to 5 (strongly agree) for the questionnaire survey done after each user study. Questions related to the same user study are plotted in the same color.

ory lapse tends to occur for asymmetric bimanual gestures; 3) Slight handle bar wobble could occur during the hand gesture change due to the shift in the computed hand centroid; and 4) Inter-hand occlusion may occur during the rotation.

APPLICATION EXAMPLES

The proposed interaction designs based on our handle bar metaphor were applied to three different applications to illustrate their potential.

The first application example shows how furniture can be arranged to a desired layout in a 3D virtual environment (see Figure 16 (top left)). The multiple object manipulation technique was used to quickly arrange similar chairs. The translate-rotate manipulation was used to "pick up" a toppled flower pot and place it on the table in one continuous bimanual hand movement. Once on the table, constrained rotation was invoked to continuously rotate the pot till it was deemed to be at the desired orientation (see video).

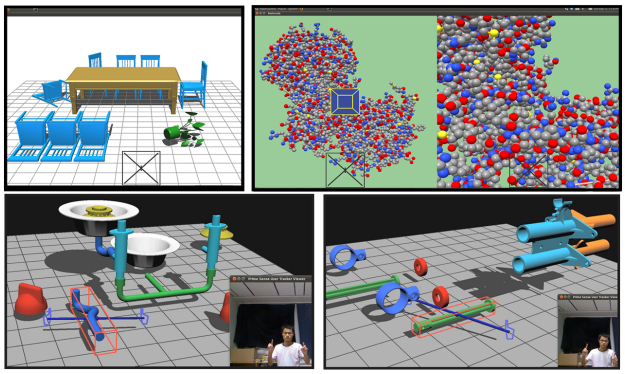


Figure 16. Applications using the handle bar manipulation techniques. Top left: Arrangement of virtual furniture in a room. Top right: Visual exploration of a complex molecular structure. Bottom: Assembling mechanical parts in CAD models.

The second application example (see Figure 16 (top right)) illustrates how the handle bar interaction designs can be applied to manipulate the virtual camera (see Figure 6 (d)) to facilitate visual exploration of a complex molecular structure. The translate-rotate manipulation allows us to visually navigate within the virtual molecular structure. The scaling gesture, when applied to a virtual camera, enables us to zoom in and out while navigating freely within the virtual 3D environment. Constrained rotation applied to the virtual camera allows us to continually scan (panning) the view around using a "cranking" gesture (see video).

The last application example (see Figure 16 (bottom)) presents how the handle bar metaphor can be used to manipulate and assemble 3D mechanical parts. Two different computer-aided-design (CAD) models, *double-range burner* and *launcher*, (bottom left and right of Figure 16, respectively) are employed here. Using our interaction methods, we can efficiently assemble the models with bare hands (see video).

CONCLUSION

We propose the handle bar metaphor as an effective way to perform mid-air interactions that manipulate the pose and scale of 3D virtual objects, suitable for use with a low-cost depth sensing device like *Kinect* in a large-display setting. The main strength of this metaphor is the physical familiarity it provides users with, as they mentally map their bimanual hand gestures to manipulation operations such as translation and rotation in the virtual 3D environment. The provision of visual cues in the form of the instantaneous orientation of the protruding virtual handle bar that corresponds interactively to the ever-changing positions of the user's two hands was observed to be very effective in providing a strong sense of control to the user during interactive visual manipulation. In addition, the flexibility and variety of interaction designs based on the handle bar metaphor have been demonstrated. These include the constrained rotation operation based on a novel "cranking" bimanual gesture and speedy techniques to manipulate and align multiple objects along a straight line using a simple combination of CLOSE and OPEN hand gestures. The virtual molecule exploration application example suggests that the same handle bar metaphor could be applied to manipulate a virtual camera to support an intuitive and flexible means of performing interactive visual navigation in a 3D virtual environment.

Observations from user studies suggest that the competency in using mid-air interaction techniques for visual manipulation is not universally innate to all users. However, interaction based on the handle bar metaphor seems to provide an intuitive way for users to quickly learn how to map the action of their bimanual hand gestures to corresponding visual manipulation tasks in a 3D virtual environment. Practice was observed to quickly improve everybody's performance and reduce the differences in skill levels among first time users. However, the issue of fast fatigue onset is still a perennial problem when using mid-air interaction for precise control.

Acknowledgment. We thank anonymous reviewers for their valuable comments, Robert Y. Wang and Sylvain Paris for sharing the CAD models, and the funding agency (A*Star SERC grant No. 092 101 0063) for the support.

REFERENCES

1. MIT Kinect Demo. `www.ros.org/wiki/mit-ros-pkg/KinectDemos`.

2. Kinect Demo Gallery, 2010. `www.openni.org/gallery` and `openkinect.org/wiki/Gallery`.

3. Q.-Z. Ang, B. Horan, Z. Najdovski, and S. Nahavandi. Grasping virtual objects with multi-point haptics. In *IEEE VR*, pages 189–190, 2011.
4. M. Annett, T. Grossman, D. Wigdor, and G. Fitzmaurice. Medusa: a proximity-aware multi-touch tabletop. UIST, pages 337–346, 2011.
5. T. Baudel and M. Beaudouin-Lafon. CHARADE: remote control of objects using free-hand gestures. *ACM Communication*, 36(7):28–35, July 1993.
6. H. Benko and A. D. Wilson. DepthTouch: Using depth-sensing camera to enable freehand interactions on and above the interactive surface. Technical report, 2009. Tech. Report MSR-TR-2009-23.
7. H. Benko and A. D. Wilson. Multi-point interactions with immersive omnidirectional visualizations in a dome. In *ITS*, pages 19–28, 2010.
8. F. Bettio, A. Giachetti, E. Gobbetti, F. Marton, and G. Pintore. A practical vision based approach to unencumbered direct spatial manipulation in virtual worlds. In *Eurographics Italian Chapter Conference*, pages 145–150, 2007.
9. D. A. Bowman, E. Kruijff, J. J. LaViola, and I. Poupyrev. *3D User Interfaces: Theory and Practice*. Addison-Wesley, 2004.
10. P. Brandl, C. Forlines, D. Wigdor, M. Haller, and C. Shen. Combining and measuring the benefits of bimanual pen and direct-touch interaction on horizontal interfaces. In *AVI*, pages 154–161, 2008.
11. CEA LIST Interactive Physics Engine Demo, 2011. `www.youtube.com/watch?v=7-HOc696g6s`.
12. L. D. Cutler, B. Fröhlich, and P. Hanrahan. Two-handed direct manipulation on the responsive workbench. In *I3D*, pages 107–114, 1997.
13. T. Duval, A. Lécuyer, and S. Thomas. SkeweR: a 3D interaction technique for 2-user collaborative manipulation of objects in virtual environments. In *3D User Interfaces*, pages 69–72, 2006.
14. W. Garage and the Stanford Artificial Intelligence Laboratory. Robot Operating System (ROS), 2009. `www.ros.org/wiki/`.
15. T. Grossman, D. Wigdor, and R. Balakrishnan. Multi-finger gestural interaction with 3D volumetric displays. In *UIST*, pages 61–70, 2004.
16. Y. Guiard. Asymmetric division of labor in human skilled bimanual action: The kinematic chain as a model. *Jour. of Motor Behavior*, 19(4):486–517, 1987.
17. G. Hackenberg, R. McCall, and W. Broll. Lightweight palm and finger tracking for real-time 3D gesture control. In *IEEE VR*, pages 19–26, 2011.
18. O. Hilliges, S. Izadi, A. D. Wilson, S. Hodges, A. Garcia-Mendoza, and A. Butz. Interactions in the air: Adding further depth to interactive tabletops. In *UIST*, pages 139–148, 2009.
19. K. Hinckley, R. Pausch, D. Proffitt, and N. F. Kassell. Two-handed virtual manipulation. *ACM Trans. on Computer-Human Interaction*, 5(3):260–302, 1998.
20. J. Jacobs and B. Froehlich. A soft hand model for physically-based manipulation of virtual objects. In *IEEE VR*, pages 11–18, 2011.
21. C. John, U. Schwanecke, and H. Regenbrecht. Real-time volumetric reconstruction and tracking of hands and face as a user interface for virtual environments. In *IEEE VR*, pages 241–242, 2009.
22. R. Jota and H. Benko. Constructing virtual 3D models with physical building blocks. In *CHI Extended Abstracts*, pages 2173–2178, 2011.
23. S. Kolarić, A. Raposo, and M. Gattass. Direct 3D manipulation using vision-based recognition of uninstrumented hands. In *X Symposium on Virtual and Augmented Reality*, pages 212–220, 2008.
24. J.-C. Lévesque, D. Laurendeau, and M. Mokhtari. Bimanual gestural interface for virtual environments. In *IEEE VR*, pages 223–224, 2011.
25. X. Luo and R. V. Kenyon. Scalable vision-based gesture interaction for cluster-driven high resolution display systems. In *IEEE VR*, pages 231–232, 2009.
26. Microsoft. Kinect for xbox360, 2010. `www.xbox.com/en-US/kinect`.
27. M. Nancel, J. Wagner, E. Pietriga, O. Chapuis, and W. Mackay. Mid-air pan-and-zoom on wall-sized displays. In *CHI*, pages 177–186, 2011.
28. R. G. O'Hagan, A. Zelinsky, and S. Rougeaux. Visual gesture interfaces for virtual environments. *Interacting with Computers*, 14(1):231–250, 2002.
29. PrimeSense, G. Willow, Side kick, and ASUS. OpenNI. `www.openni.org`.
30. Y. Sato, M. Saito, and H. Koike. Real-time input of 3D pose and gestures of a user's hand and its applications for HCI. In *IEEE VR*, pages 79–86, 2001.
31. J. Segen and S. Kumar. Gesture VR: vision-based 3D hand interface for spatial interaction. In *ACM Multimedia*, pages 455–464, 1998.
32. R. Y. Wang, S. Paris, and J. Popović. 6D hands: Markerless hand-tracking for computer aided design. In *UIST*, pages 549–558, 2011.
33. B. Yoo, J.-J. Han, C. Choi, K. Yi, S. Suh, D. Park, and C. Kim. 3D user interface combining gaze and hand gestures for large-scale display. In *CHI*, pages 3709–3714, 2010.
34. J. Zigelbaum, A. Browning, D. Leithinger, O. Bau, and H. Ishii. g-stalt: a chirocentric, spatiotemporal, and telekinetic gestural interface. In *Intl. Conf. on Tangible and Embedded Interaction*, pages 261–264, 2010.