www.cadanda.com Computer-Aided Design AND APPLICATIONS

An Intuitive Human-Computer Interface for Large Display Virtual Reality Applications

X.B. Yang[1], S. H. Choi[2], K. K. Yuen[3] and L. K. Y. Chan[4]

The University of Hong Kong
[1] xbyang@hku.hk, [2] shchoi@hku.hk, [3] ryanyuen@hku.hk, [4] lkychan@hku.hk

ABSTRACT

The paper presents a human-computer interface (HCI) for natural interaction in the Cave Automatic Virtual Environment (CAVE). By tracking hand actions and movements with a set of six infra-red (IR) cameras, the proposed HCI allows a user to interact directly with virtual objects in the CAVE without the need to attach any complex electronic tracking devices, such as data-gloves, on the user's hands. An intuitive control scheme based on simple and natural hand gestures, including "grab" and "drop", is developed to demonstrate the strengths of the HCI. A case study of using the HCI to facilitate the simulation of fork-lift truck operations in the CAVE system is presented. Finally, the compatibility and benefits of integrating the HCI with virtual reality (VR) simulation systems in the CAVE are discussed.

Keywords: virtual environment interaction, hand tracking, large display VR, CAVE.
DOI: 10.3722/cadaps.2010.269-278

1 INTRODUCTION

Despite the massive popularity of virtual reality (VR) in sci-fi media and the general portrait of its being a symbol of the future world, some are dubious about VR due to its complex human-computer interface (HCI) and poor interaction quality. Thomson [17] pointed out that VR is a technology that "has promised so much and delivered so little". Over the years, researchers have spent huge efforts to develop the ultimate user interface (UI) suitable for VR applications, from classic desktop input devices to biological tracking devices like motion tracking, voice recognition or even brain-computer interfaces, such as "The Berlin Brain-Computer Interface (BBCI)" [12]. However, these technologies are still far from being able to impact ordinary users with the similar "Wow" effect like when they saw Tom Cruise tackling crime cases in the "Minority Report" movie. An apparent paradox seems to be the creation of an immersive interface where the user is part of the virtual 3D world while not constraining his or her movements with clumsy and awkward-looking devices. This dilemma aggravates in large display environments, such as the CAVE.

Although improvements in visual aspect have relieved users from heavy and uncomfortable head-mounted displays (HMD), the input paradigm has been stuck with traditional device-based controls ever since the creation of the world's first CAVE [4]. Recently, researchers have endeavoured to develop friendly HCIs for large-display VR applications, making much contribution to device-based and gesture-based interactions, as follows:

- **HCI for device-based interaction:**

  These HCIs are basically enhancements of the classic 2D touchpad. Rekimoto [15] introduced a natural "Pick-and-Drop" gesture into the multi-user collaborative working environment which relies on interactions of the digital whiteboard and hand-held devices. Hachet's "Interaction Table" [8] also focuses on building a stable multi-user interface for large display, which is developed on an orientation-tracking 2D touchpad. Grossman [7] designed a unique hand-held device for two-handed control applied in 3D modelling on a large projection display. Malik [13] added a 3D tracker to an optical touchpad, which relies on camera captured images of the user's fingers for interaction. Cao [2] improved the traditional pointing device into the track-able "VisionWand" to carry out 3D manipulation commands. Bowman [1] reviewed several customized input devices such as the "Finger Sleeve", "Cave Painting", "Interaction Slippers" and other glove-alike accessories, for application in CAVE. While the "Finger Sleeve" serves as a portable version of the traditional 3D mice, the latter three provide natural control in the virtual environment (VE).

  The benefits of applying device-based interaction, either in 2D or 3D, are obvious: devices can be specially optimized for applications in a certain VE. For touchpad interaction, the technology is mature and has been commercialized. It can thus support stable and accurate control. For those who adopt joystick control in VR simulations, the learning period of first time users can be significantly shortened since people are already familiar with the 2D input gadgets.

  However, a big limitation of the device-based HCI is that the machine between human and virtual objects serves not only as a translator of the user's ideas, but also as a wall which blocks the user from direct interaction with the virtual world. Moreover, these devices constrain the user's movements. For most cases, instead of interacting with virtual objects intuitively, the user has to memorize shortcut keys or commands to accomplish complex tasks.

- **HCI for gesture-based interaction:**

  Gesture is the major language in this research area. Vogel [18] developed a basic freehand control mechanism for large 2D display applications, in which the user stands at a distance from the screen. Kim [11], on the other hand, turned the projection walls into touch screens through IR tracking. Oblong Industries' "g-speak" [5] is a commercialized HCI considered as the symbol of all current gesture-based interaction: it uses IR cameras for motion and gesture tracking, translates hand or body gestures into commands, while the user interacts with a single-sided display wall. Cipolla and Okamoto [3] also introduced an algorithm for hand gesture motion tracking that can be translated into 3D model control commands.

  This HCI gets rid of complex attachments on the user, who can focus on the creation of basic UI instead of on specific applications. As listed by Graham-Rowe [6], gestural control is no doubt among the best computer interfaces. Nevertheless, waving or posing in front of a big screen cut the user out of virtual world, and those complex gestures may be neither natural nor friendly for first time users.

This paper therefore proposes an intuitive HCI for VR users "to move around in the virtual world and see it from different angles, to reach into it, grab it, and reshape it." [16] The proposed HCI framework focuses on natural user experience in CAVE. It is characterized by the following aspects:

- Direct interaction between the user and the virtual world;
- A natural, intuitive control scheme, with simple UI that mimics the way humans interact with the real world;
- No electrically triggered devices attached, and no complex gestural commands;
- Relatively low cost, easy to implement, and can be seamlessly integrated with VR systems. The IR camera motion tracking required is now a standard feature of modern CAVE. Our control scheme uses only basic motion tracking that can be accomplished at a cost of about US$5,000; this is inexpensive compared to other high accuracy IR tracking solutions.

2 CONCEPTUAL DESIGN AND IMPLEMENTATION

2.1 IR Camera Hand Motion Capture

An IR camera can capture live images of near infra-red spectrum, as ordinary video cameras do with visible spectrum. When lit up by arrays of IR LEDs, objects with IR reflective coatings will be captured by IR cameras as vivid markers differentiated from the rest of the image. In the project, OptiTrack™ [14] IR motion capture system is integrated to capture hand movements using this mechanism, the software Tracking Tools™ was selected for motion tracking and gesture recognition. In the prototype design, three segments of reflective tape are stuck to the user's thumb finger tip, middle finger tip, and palm's edge, where they will be captured as vivid markers. In Tracking Tools™, which is the software tracking solution of OptiTrack™, the 2D positions of the markers captured by each IR camera are combined and calibrated to reconstruct the actual 3D locations. These markers are then visualized in the Tracking Tools™ as virtual trackers; by selecting three or more virtual trackers to form a certain shape, such as the triangle in Fig. 1, the shape will be recognized as a "rigid-body" for motion tracking of the object with those markers attached on. Later in the development, we used thin black gloves sewed with the 3M Magic Tape™ to make it easier to shift the positions of trackers for tune-ups and experiments.

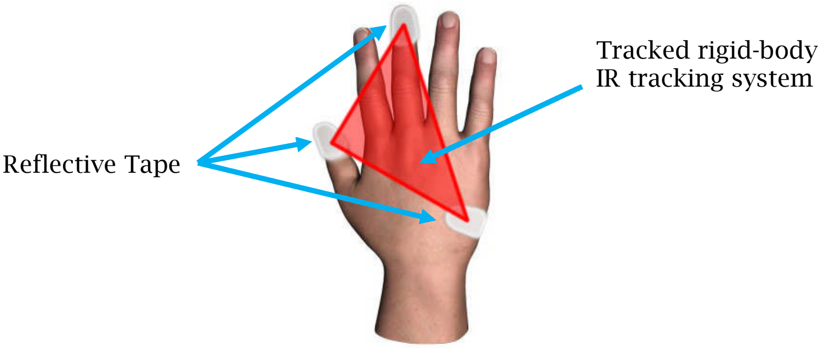


Fig. 1: Hand tracking prototype design.

2.2 The imseCAVE

Fig. 2 shows the schematic layout of the imseCAVE [9]. It is an in-house, full-immersive VE developed by the Department of Industrial and Manufacturing Systems Engineering at the University of Hong Kong. It is about 3m × 3m × 2.5m, and is based on the technologies of distributed VR.

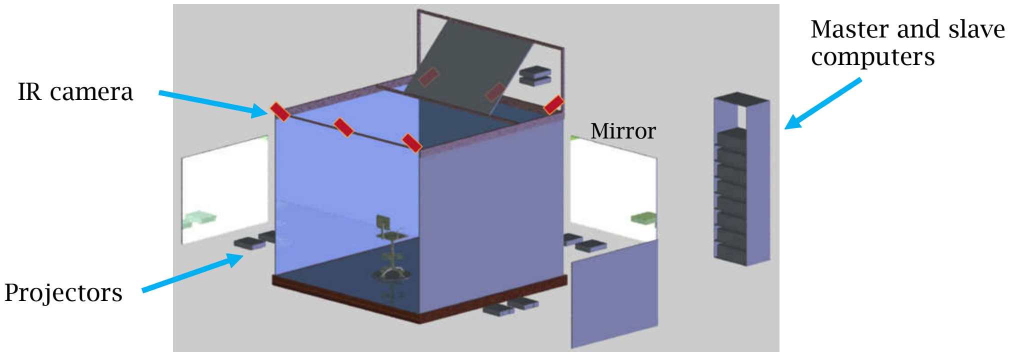


Fig. 2: Schematic layout of the imseCAVE with 6 IR cameras.

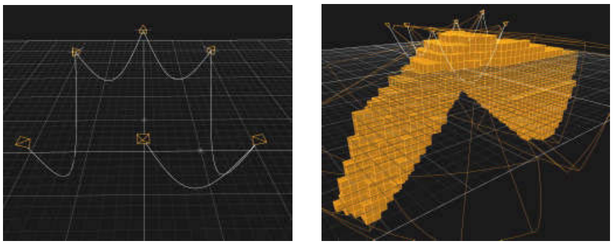


Fig. 3: The cameras' locations and tracking coverage in Tracking Tools™.

By adding a set of six IR cameras, we can track a user's position and movements in the imseCAVE and transmit the signals into the virtual world. By tweaking the cameras' locations and settings, more than 70% of the imseCAVE space can now be covered. Fig. 3 shows the cameras' locations and tracking coverage recorded in Tracking Tools™. Each IR camera can track up to 100 frames per second with a resolution of 640 × 480; the latency is less than 5ms and the tracking error is within mm scale.

2.3 The "Grab 'n Drop" Control Scheme

To capture a user's hand gestures, we sample and define two different shapes of the same three tracker balls when the user makes grabbing and dropping hand gestures. The three tracker balls are comprised of one dedicated tracker ball on the middle finger and two of the three tracker balls that define the hand's 3D position, presented by the yellow triangle in Fig. 4, where the red and green ones show different shapes when the user makes "Drop" or "Grab" gestures.

Fig. 4: Grab and drop gestures and corresponding tracked shapes.

We improved the prototype design for more precise and faster motion tracking. In the original design, we used the "drop" shape for hand motion tracking. Later in the implementation, this scheme was found to demand too much steadiness of a hand gesture; for constant motion tracking, the trackers would need to stay still relative to each other, demanding the user to maintain a certain hand shape while moving around, which is unnatural and inconvenient. Therefore we designed a gadget with three tracker balls attached on and held relatively still to form a static shape for hand motion tracking. Two of the tracker balls are shared by a fourth tracker ball for tracking hand gestures. To pick up something, the user's hand forms the "Grab" shape, while to put down the object, the fingers loosen up. Consequently, the shape formed by the relative positions of the tracker balls changes from "Grab" to "Drop".

Tracking Tools™ broadcasts the captured information under VRPN protocol through a local LAN connection, and Virtools™ (the software kernel for virtual world construction and control of the imseCAVE) translates the orientation and position data into the virtual world. The hand trackers are initialized as a virtual objects, and the "Grab", "Drop" shapes serve as triggers that set a binary state variable to true or false. The user's position inside the imseCAVE is captured by tracking the tracker balls attached to the stereo-glasses that the user wears. As such, we align the user's eyes and hands in the virtual world, in which the user can directly interact with the virtual objects.

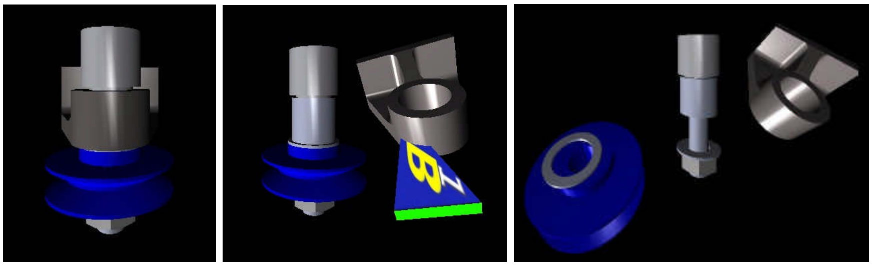
3 APPLICATIONS OF THE PROPOSED HCI

To demonstrate its strengths, the proposed HCI has been implemented in several simple applications for simulating different situations, including virtual assembly, VR navigation, manipulation, and game interface design.

3.1 Virtual Assembly

The idea of virtual assembly basically relies on collision detection of user's hand and action detection of user's pick-up, put-down gestures. As shown in Fig. 5, the blue triangle marked with "1B" indicates user's hand position and orientation in the virtual world. When the triangle is in contact with a part and the user makes a grab action, the part will stick to the user's hand to simulate the "pick-up" action. Wearing a pair of stereo-glasses in the imseCAVE, the user will actually see this part being "picked up" and "held" in the hand. To "put-down" the part, the user simply loosens his or her fingers to trigger a "drop" action.

Virtual assembly has often been awkward on classic 2D input devices. Although data-gloves can mimic natural interaction, they can very much constrain the user's movements. Our solution strikes a balance between precision and intuition. It provides fast, seamless and natural control. More importantly, when the user is in the imseCAVE, the blue triangle will be replaced by the user's hand. Therefore instead of feeling like using a pointing device in front of a giant computer screen, the user can directly "touch", "pick up" and "move around" virtual objects. Audio feedback can also be added when the user triggers those actions to enhance the experience. Currently we are working on designing proper visual feedbacks for different actions. Furthermore, we are going to integrate the physics engine provided by Virtools™ to develop more realistic simulations.



(a) Original Pulley (b) Picking up a part (c) Final result

Fig. 5: A virtual assembly process.

3.2 Solving the Rubik's Cube

In this application, we introduced the double-handed control, navigation, and accurate manipulation of objects in the imseCAVE. The user wears a pair of gloves with tracker balls that form two different shapes on the left hand and the right hand, as shown in Fig. 6, to be recognized as two different objects by Tracking Tools™.

Fig. 6: A pair of gloves for double-handed control.

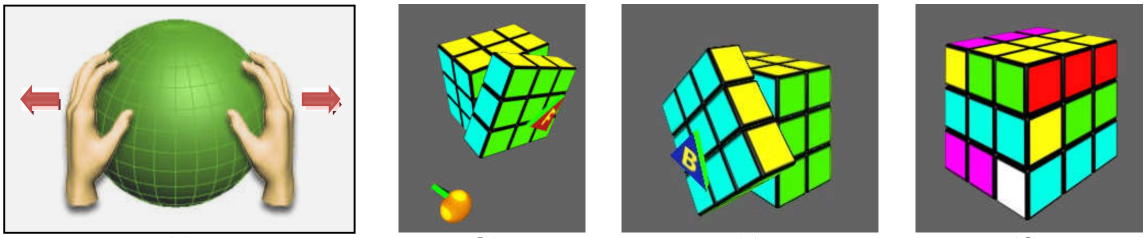
We designed two navigation schemes for this case:

- Virtual joystick:

  When the user closes in both hands, makes a “Grab” gesture and “pulls” up his or her right hand with the left hand staying still, a virtual joystick will form between the palms. The virtual joystick can be used to navigate through the virtual world, as we do when using a real one; the control scheme is identical, but more flexible as the translate accuracy can be altered by adjusting the distance between two hands. To deactivate the virtual stick, the user can simply loosen up the right hand fingers and it will disappear immediately. The virtual joystick is designed for fast navigation, which can be very handy for applications like 3D map surfing.
- “Grab ‘n Drop”:

  After navigating to the target area, the user’s view angle or position can be fine-tuned in an intuitive manner called “Grab ‘n Drop”. The user can “grab” the space around a virtual object and move the whole virtual world (including the object) about, then “drop” it wherever he or she wants, similar to the way people approach an object on a dining table either by pulling the table cloth towards them, or instead by leaning towards the object. The leaning approach is simulated through head tracking in this case.

Manipulation of virtual objects is also accomplished through cooperative actions of two hands. If the user closes in both hands and “pull” up vertically as in the previous case, the virtual joystick will be activated. But if he or she “pulls” horizontally, a manipulation sphere appears and sticks to the left hand. Afterwards, the object’s position or orientation can be adjusted by moving around or rotating the “Manip-Sphere”. The user can close in hands again to deactivate the “Manip-Sphere”.



(a) (b) (c) (d)

Fig. 7: (a) Activation of “Manip-Sphere”; (b) Rubik’s cube being manipulated by the orange “Manip-Sphere” at the right bottom with the right hand rotating one side of the cube; (c) Solving puzzle; (d) Manipulation result.

Now to solve the Rubik’s cube puzzle, the user first navigates to a comfortable view angle and uses his or her right hand to “grab” and “rotate” the side, then rolls the cube over to view and work on other sides by operating the “Manip-Sphere” on the left hand. During the rotation process, the system will monitor the rotation angle. If the angle is over 45^{o}, the side will automatically rotate to the full 90^{o}. As such, the user will not have to twist his or her wrist just to rotate one side of the cube.

In comparison with the traditional simulated Rubik's cube puzzles which generally rely on indirect 2D controls, our puzzle can be solved in a way humans do in the real world. This shows the strength of our HCI for fast navigation and precise manipulation, and its possible value for surfing 3D maps and architectural modelling.

3.3 Virtual Archery

This application is designed to show the potential of "Grab 'n Drop" interface in game industry. First-person-shooter (FPS) games are considered applications of the non-immersive VR, and there have been efforts in porting FPS games or game engines into CAVE [10]. For our case, we have created a bare-handed control scheme for shooting games.

As shown in the Fig. 8, archers assume similar poses to shoot, in that they hold the bow in left hand, grab the string near the left hand with an arrow in the right hand and pull back to draw the bow, aim by adjusting the relative position of left and right hand, and finally loosen the right hand fingers to shoot.

Fig. 8: Archers posing to shoot (Courtesy of Wikipedia).

With the same "Grab 'n Drop" interface, we can mimic the real life archery process above in the virtual world: a user standing in the imseCAVE, with a virtual bow in his or her left hand, grabbing the "string" with right hand near the left hand to draw, adjusting the position of both hands to determine the direction, and finally releasing the right hand fingers to shoot. Indeed, this control scheme gives the user more flexibility when interacting with the virtual world and at the same time provides full immersive experiences. This archery process is demonstrated in Fig. 9, in which the player is wearing a pair of stereo-glasses with tracker balls to indicate his position in the imseCAVE. The overlapping images create a stereoscopic illusion that the bow and arrow are in the user's hands.

(a) Grab (b) Draw (c) Aim (player's view) (d) Shoot

Fig. 9: Virtual archery.

4 CASE STUDY: INTEGRATION OF THE HCI WITH AN ON-GOING PROJECT

This section describes a case study of integrating the HCI interface with an on-going VR simulation project to further highlight its flexibility. The integration was completed in only three days, from design to successful implementation.

4.1 Background

A VR system for simulation forklift truck operations in the imseCAVE, as shown in Fig. 10, is being developed in the Department. It is built to simulate possible accidents to help enhance safety mentality of forklift truck drivers. In this system, force feedback wheel with gas and brake pedals is used to control the truck, while Virtools™ is adopted for 3D rendering and physics simulations.

The purpose of integrating the HCI interface is to improve controllability of the forklift truck, and to change several parameters in real-time to simulate different operations, which can sometimes be quite cumbersome before implementation. To alleviate this problem, we first added head tracking to the VR system to enhance the user's experience when "driving" in the imseCAVE. Then we considered adding a flexible menu based on the "Grab 'n Drop" interface for intuitive and convenient operations, instead of pressing the keyboard or setting up a couple of shortcut keys on the wheel. The desired menu functions are:

- Altering scenes with different terrains;
- Changing the loading's weight in real-time;
- Switching camera among the driving view, chase view and free navigation view;
- System functions: reload, reset and exit.

Fig. 10: The forklift truck simulation.

4.2 Menu and Control Scheme Design

We expanded the menu functions above into a curved four-section 3D menu, as shown in Fig. 11. The elements of this menu are differentiated by four colours, so that the operator can easily identify when browsing through the sections. When activated, the menu will surround the operator in the front centre, called the "active zone". As one section of elements is highlighted, all other sections gradually fade out.

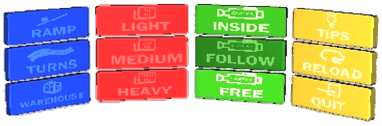
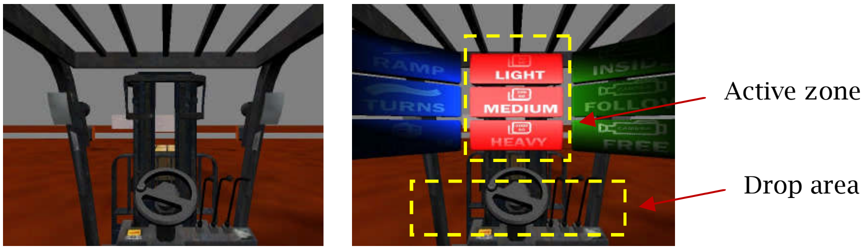


Fig. 11: The menu design.



(a) (b)

Fig. 12: Driver's view: (a) Before menu implementation, (b) After menu implementation.

To active the menu, the driver simply reaches out his or her right hand in the up-right direction, makes a “grab” and “pull” action as if dragging a scroll on the windshield of the simulated truck. The menu will follow the driver’s hand and slide smoothly into the centre. To select an item on the menu, the driver just “grabs”, “pulls” down and “drops” the item in the “drop area”. To switch to other sections, the driver only needs to “grab” anywhere outside of the “active zone”, and drags the menu around to the left or to the right. The menu will automatically retreat to the up-right side if any item is dropped in the “drop area” or there is no activity captured from the driver.

4.3 Implementation

The driver needs to wear a glove with tracker balls, as shown in Fig. 13, to simulate driving the imseCAVE. The glove is the only extra investment on the hardware part of the VR simulation system. The rest of the HCI integration is mainly software implementation in Virtools™. It only took one day to port the HCI codes and add collision rules for menu interactions. The result is shown in Fig. 14.

Fig. 13: Driver wearing a glove with tracker balls.

(a) Normal driving (b) Activating (c) Browsing

Fig. 14: Interacting with the menu for forklift driving.

5 CONCLUSIONS AND FURTHER DEVELOPMENT

In this paper, we proposed an intuitive human-computer interface (HCI) for natural interaction with large display VR applications in the CAVE. In comparison with other works, the major strengths of the proposed HCI include:

- Flexibility – easy to customize and implement, device independent;
- Compatibility – a general interface framework capable of major types of VR interaction: navigation, manipulation, menu operating, etc.;
- User-friendliness – direct, natural and intuitive control.

The HCI was developed with a main emphasis on the user-oriented interface. Instead of requesting the user to learn all sorts of complicated and confusing devices, we aimed to incorporate natural human behaviours in the HCI as the ultimate goal of VR research. Therefore, developing software systems that can automatically sample and learn different users’ gestures to facilitate VR applications remains paramount.

To take advantage of and further develop the HCI, we aim to build an integrated development environment (IDE) in the imseCAVE for product design and manufacture, including CAD modelling, virtual prototyping, product testing and evaluation. As such, designers can directly manipulate product prototypes, try out modifications, and make design changes easily. The IDE indeed provides a highly efficient and cooperative environment in which people with different expertises can effectively share their ideas, and thus benefit from seamless collaborations. As a result, we can shorten the current product design and development circles hugely.

REFERENCES

[1] Bowman, D. A.; Kruijff, E.; LaViola, Jr., J. J.; Poupyrev, I.: 3D user interfaces: theory and practice, Addison-Wesley, Boston, 2005.

[2] Cao, X.; Balakrishnan, R.: VisionWand: interaction techniques for large displays using a passive wand tracked in 3D, Proceedings of ACM SIGGRAPH 2004, ACM, New York, NY, USA, 2004, 729.

[3] Cipolla, R.; Okamoto, Y.; Kuno, Y.: Robust Structure from Motion using Motion Parallax, Proceedings of the IEEE 4th International Conference on Computer Vision, Berlin, Germany, 1993, 374-382.

[4] Cruz-Neira, C.; Sandin, D. J.; DeFanti, T. A.: Surround-screen projection-based virtual reality: the design and implementation of the CAVE, Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, ACM, New York, NY, USA, 1993, 135-42.

[5] g-speak spatial operating environment, http://oblong.com/, Oblong Industries.

[6] Graham-Rowe, D: The Best Computer Interfaces: Past, Present, and Future, http://www.technologyreview.com/computing/22393/, MIT Technology Review, 2009.

[7] Grossman, T.; Balakrishnan, R.; Kurtenbach, G.; Fitzmaurice, G.; Khan, A.; Buxton, B.: Interaction techniques for 3D modelling on large displays, Proceedings of the 2001 symposium on Interactive 3D graphics, ACM, New York, NY, USA, 2001, 17-23.

[8] Hachet, M.; Guitton, P.: The interaction table: a new input device designed for interaction in immersive large display environments, Proceedings of the workshop on Virtual environments 2002, Eurographics Association, Aire-la-Ville, Switzerland, Switzerland, 2002, 189-96.

[9] imseCAVE, http://www.imse.hku.hk/intellisyslab/facilities/imseCAVE.htm, Industrial and Manufacturing System Engineering Department, The University of Hong Kong, Hong Kong.

[10] Jacobson, J.; Le Renard, M.; Lugrin, J.-L.; Cavazza, M.: The CaveUT system: immersive entertainment based on a game engine, Proceedings of the 2005 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology, ACM, New York, NY, USA, 2005, 184-87.

[11] Kim, H.; Fellner, D. W.: Interaction with Hand Gesture for a Back-Projection Wall, Proceedings of the Computer Graphics International, IEEE Computer Society, Washington, DC, USA, 2004, 395-402.

[12] Krepki, R.; Curio, G.; Blankertz, B.; Müller, K.-R.: Berlin Brain-Computer Interface-The HCI communication channel for discovery, International Journal of Human-Computer Studies, 65(5), 2007, 460-477.

[13] Malik, S.; Ranjan, A.; Balakrishnan, R.: Interacting with large displays from a distance with vision-tracked multi-finger gestural input, ACM SIGGRAPH 2006 Sketches, ACM, New York, NY, USA, 2006.

[14] OptiTrack™, http://www.naturalpoint.com/optitrack/, NaturalPoint.

[15] Rekimoto, J.: A multiple device approach for supporting whiteboard-based interactions, Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Addison-Wesley Publishing Co., Los Angeles, CA, USA, 1998, 344-51.

[16] Rheingold, H.: Virtual Reality, Simon & Schuster, New York, 1991.

[17] Thomson, I.; Nichols, S.: Top 10 disappointing technologies, PC Authority, 2009, http://www.pcauthority.com.au/News/145271,top-10-disappointing-technologies.aspx.

[18] Vogel, D; Balakrishnan, R.: Distant freehand pointing and clicking on very large, high resolution displays, Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, ACM, New York, NY, USA, 2005, 33-42.