

A Natural User Interface for Navigating in Organized 3D Virtual Contents

Guido Maria Re and Monica Bordegoni

Politecnico di Milano, Dipartimento di Meccanica,
Via La Masa, 1, 20156 Milano, Italy
{guidomaria.re,monica.bordegoni}@polimi.it

Abstract. The research activity presented in this paper aim at extending the traditional planar navigation, which is adopted by many desktop applications for searching information, to an experience in a Virtual Reality (VR) environment. In particular, the work proposes a system that allows the user to navigate in virtual environments, in which the objects are spatially organized and sorted. The visualization of virtual object has been designed and an interaction method, based on gestures, has been proposed to trigger the navigation in the environment. The article describes the design and the development of the system, by starting from some considerations about the intuitiveness and naturalness required for a three-dimensional navigation. In addition, an initial case study has been carried out and consists in using the system in a virtual 3D catalogue of furniture.

Keywords: Virtual Reality, Natural User Interfaces, Navigation, Gestures, Virtual Catalogue.

1 Introduction

A considerable amount of time that users spend interacting with a computer is usually devoted to information search. These pieces of information are virtual contents, such as pictures, news, music and videos or, in other more articulated cases, also 3D models, and they are usually represented by means of thumbnails spatially organized. The reason behind their disposition is to make them clearly visible and easy to recognize. The disposition side by side of these elements is often carried out according to logical reasons or a classification. Typical examples are the e-commerce websites, in which goods on sale are ordered by relevance or price, or the slideshow visualization mode to select a tune to play, where the music albums are arranged in alphabetical order or by genre.

Traditional interaction devices, such as keyboard and mouse, provide the browsing experience to the user. The navigation is performed by a scrollbar or other dragging interaction metaphors that limit the user's experience in one direction. Hence, the navigation through these pieces of information is limited along a single direction, which could be the vertical or the horizontal one.

From these considerations, it turns out that searching information is a common aspect for many traditional computer applications and websites. However, the visualization

and interaction methods for these purposes limit the user navigation experience. They are actually oriented to desktop-like user interfaces. The effectiveness of these navigation methods dramatically decreases if used in particular HCI systems, such as in Virtual Reality (VR) environments. Therefore, new solutions for browsing in virtual spaces become necessary.

The objective of this research work is to overcome the limitations of the current navigation interfaces across organized objects in a tri-dimensional VR environment and to extend the user's navigation possibilities by means of Natural User Interfaces (NUI) [1]. In this way the user is not in a flat virtual space and he can experience a free 3D navigation.

The proposed navigation system is oriented to use spatial gestures to enable a 3D navigation experience in an organized Virtual Reality (VR) environment. The items in the VR environment are sorted according to common features. Moreover, in order to promote and expand new contexts of use for VR, the gestures must be as natural as possible and the technologies involved must be cheap. Actually, the need of interactive methods, which are easy to learn and remember, and the use of mass-market devices, which are highly available and for a low price, are the two requirements for a large-scale deployment of NUIs.

The paper is organized as follows. Section 2 gives an overview of the state of the art of the NUIs and virtual navigation systems. Section 3 describes the system while Section 4 the design of the gestures used for the interaction. Section 5 presents an initial case study. The paper closes with a discussion and conclusions.

2 Background

NUI's are computer-interaction modalities that have been growing up in the last years. The term "natural" is due to the possibility of managing a computer through usual gesture and minimally invasive devices. NUI enables the user to provide inputs to a machine by means of his whole body and movements in the space, differently to the traditional mouse and keyboard [2]. This modality is particularly convenient in wide interactive environments, such as the ones provided by a CAVE [3]. In these cases, in fact, the user is free to move and he does not have to hold any bulky device. Finally, the possibility of interacting with natural gestures makes the NUIs intuitive, easy to learn and increases the user's engagement during the use. According to these considerations, NUIs are more usable than traditional interaction devices [4].

The great progress in the miniaturization, the development of powerful technologies and efficient software brought to new NUI device available for a cheap price in the last decade. In particular, several new devices came out in the video-game mass market. These devices have been nicely integrated into research projects in the field of virtual navigation and interaction. One of the first examples was the use of the Balance Board from Nintendo¹ as input device for moving in a virtual space [5]. However, these game devices require a physical interaction between them and the user in order to provide an input to the machine.

¹ Nintendo Balance Board - <http://wiifit.com/>

Advances in Computer Vision (CV) and HCI brought also the whole human body to become an interface without wearing any device. Smart systems are currently able to recognize a user by means of cameras [6], and interpret his/her movements as computer inputs avoiding the touch interaction. [7] evaluated a simple system, based on a single camera, to detect the hands and the head of a user in a CAVE for navigation purposes. More recently, the release of Microsoft Kinect for Xbox 360² pushed further the boundaries of NUI devices by providing a cheap motion capture system. Kinect couples the video stream of a normal RGB camera with depth information and it is also able to detect a user and the position of 21 joints of his/her body in real time. [8] proposed to use this device for map navigation purposes by detecting hand direction and motion by means of the depth data. An important contribution on the spread of Kinect has been given by the middleware called FFAST [9], which is able to recognize body gestures and consequently trigger keyboard events. A gesture-driven navigation system for VR environments by using Kinect and FFAST is described in [10].

Differently to the previous researches, this work deals with a virtual environment where objects are precisely organized in the space according to a three-dimensional grid. The objects are sorted and arranged in the space according to classification features. For this reason, free navigation in the space is not requested because the browsing is from a precise coordinate point to another one. These two points correspond to the spatial position of two different objects. [11] proposed a similar approach to navigate through the media contents on a television by means of gestures and a camera sensor. [12] developed a system addressed to browsing a database of medical images in a hospital, with the advantage for the doctor to avoid touch interactions in a sterile environment. [13], finally, developed a similar system to browse items by using Kinect. The research described in this paper takes into account these previous works and it contributes to extend the navigation user's experience to three dimensions.

3 System Description

The system proposed provides a 3D navigation through a large display and a NUI. As depicted in Fig. 1, the system is made up of hardware and software components that work all together to allow the user to experience the navigation. The components are organized in two main modules according to their task; an event driven approach triggers the communication between the modules. The first module regards the user interaction experience and its task is to detect the user's gesture and to provide the related event. The second module provides the navigation experience to the user, according to the triggered event, by rendering the camera motion in the VR space. The components and the two modules are described in depth in the following.

² Microsoft Kinect - <http://www.xbox.com/en-US/kinect>

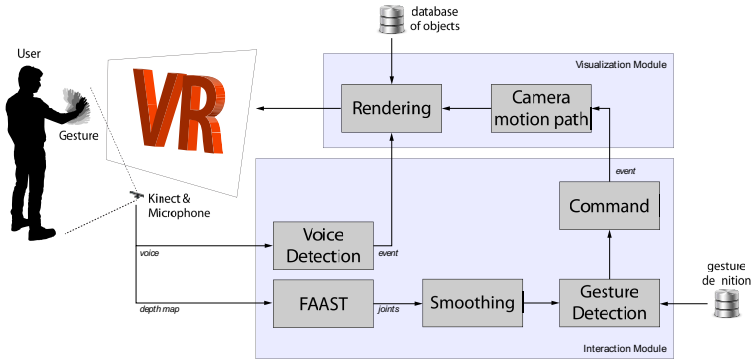


Fig. 1. Schematic representation of the system

3.1 System Components

The main hardware components of the system are the Microsoft Kinect and a PowerWall. Kinect is a motion capture controller that integrates a RGB camera, a depth sensor and an array of microphones. The color camera has 8-bit VGA resolution (640 x 480 pixels) working at 30 Hz. The depth sensor consists of a monochrome VGA camera working in the infrared spectrum (IR) and an IR laser projector. The sensor is able to detect the 3D scene in a distance range from 0.8 to 3.5 meters. The PowerWall is a retro-reflective display, whose size is 3x2.25 meters. The image stream is provided by a color projector with a resolution of 1400 x 1050 at 80 Hz. All the devices are connected to a laptop, which manages all the data in order to elaborate the navigation experience. The laptop is equipped with a 2.66 GHz processor, 4 GB RAM and a NVIDIA GeForce GT 330M GPU.

3.2 Interaction Module

The purpose of the interaction module is to trigger the navigation through the recognition of user's gestures. In order to do this, the depth map generated by Kinect is elaborated in real time by FFAST. As in Fig. 2, FFAST extracts the user shape in the scene and calculates a 3D skeleton model of the user. This skeleton is a simple representation of the biomechanical structure of the whole user body in the space and it is made up of 24 joints. Then, FFAST makes the position of each joint available to other applications through a VRPN communication network [14].

The command recognition software collects the VRPN data of the joints and works to define in real-time what the user wants to command through gestures. This software has been developed by means of Virtools 5.0³, which is a game engine by Dassault Systèmes. As a first step, the command recognition software filters the received data along time, in order to remove high frequency noise. This step is carried out by means of a low-pass filter. In this way, the system makes smooth the raw user's

³ 3DVIA Virtools - <http://www.3ds.com/products-services/3dvia/3dvia-virtools/>

movements captured by Kinect. Subsequently, the command recognition software uses the data and detects precise gestures according to information stored in a database. Since the navigation is in a 3D organized space, the system requires six different gestures to move in all of the main directions. The system compares the user movements with the gestures described in the database and an event is generated in order to trigger the navigation in case of matching. The definitions of the gestures and the detection algorithm are reported in Section 4 in detail.

Lastly, additional events are generated by voice commands. A microphone placed in the environment captures the user voice and detects precise words. In particular, the two words are “select” and “release”. The system recognizes these two words and allows the user to pick or leave a virtual object in the 3D environment.



Fig. 2. The visualization of the depth map from Kinect by means of FFAST. The light blue area corresponds to the user, while the red lines indicate his simplified skeleton.

3.3 Visualization Module

The Visualization Module enables the user to see the virtual space and to browse across the objects, which are arranged in a particular way. The objects are placed on fixed positions in the space so that they are at the same distance to each other. In addition, these objects are located in the virtual space according to three classifications. Actually, each object can be described according to some qualitative and quantitative features. The three common features among the objects are associated to the three dimensions of the environment. Consequently, they can be arranged in a sort of visual 3D array, where each place corresponds to a value or a quality of a feature. Figure 3 represents an example of such arrangement by means of coloured primitive solids. Different kinds of objects are arranged in the space according three features:

- Shape, (cone, pyramid, cube, cylinder, sphere)
- Size, (extra small, small, medium, large, extra large)
- Colour (red, orange, yellow, green, light blue, blue, purple)

Thus, each feature is associated with a dimension and the navigation along one direction allows the user to browse through the objects with different values associated with one of the features. In this particular example, the user switches to a different colour by moving horizontally, changes the object dimension going forward and backward, selects another shape going up and down.

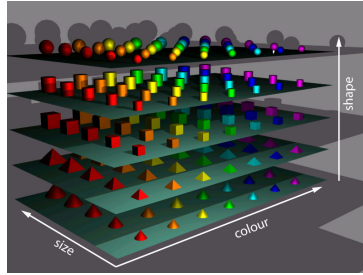


Fig. 3. Several 3D objects are arranged in the space according to the shape, size and colour

The navigation through these objects is performed by means of the motion of the point of view of the virtual camera on a precise path. The path is a straight line along one direction and it connects one object with the neighbouring one. The total motion time to switch from two objects has been set to 0.8 seconds.

In particular cases, navigation in a direction is impossible, since the user is watching an object at the border of the environment. Thus, the system provides a visual feedback to the user, in order to alert him/her that he cannot move towards the selected direction. The feedback is a horizontal shaking motion of the camera.

4 Gesture Interface

This section describes the definition of the gestures and the implementation of the interface in detail. In particular, it focuses on the design according to the human-based approach proposed in [15]. The objective is to provide a NUI with gestures that are easy to make from an ergonomic point of view, easy to remember and logically coupled to the functions they have to perform. Therefore, six different gestures have been designed and an algorithm to detect them has been developed. Finally, useful parameters for the algorithm have been evaluated by measuring some human movements.

4.1 Design of Gestures

The function of the six gestures is to indicate the moving direction of the virtual camera to the system. All of the gestures start from the position called *neutral* in Fig. 4a. No commands are executed while the arm is in this position. The position consists of placing the right hand in front of the shoulder, by keeping the elbow relaxed and bended.

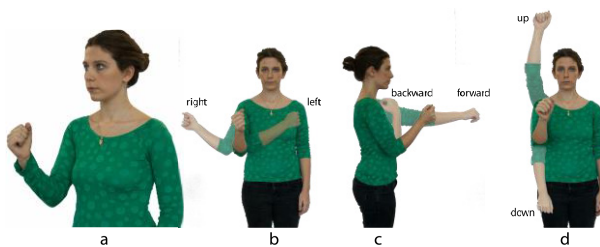


Fig. 4. The neutral position (a) and the six gestures defined for the navigation in the virtual environment (b, c and d)

The gestures involve the motion of the hand and the arm to six different positions. Abduction of the arm from the neutral position means a camera motion to the right, while an adduction on the horizontal plane means a motion to the left (Fig. 4b). Extension and contraction of the arm indicate respectively forward and backward movement (Fig. 4c). Finally, an extension of the forearm to the hip specifies a downward movement of the camera, while an upward flexion of the arm entails the camera to move above (Fig. 4d).

4.2 Gesture Detection Algorithm

The developed detection algorithm receives the position of all the user’s joints and recognizes if the user is performing one of the gestures previously defined. The data coming from FAAST are elaborated in order to estimate the position of the right hand, relative to the right shoulder, and the two angles, as shown in Fig. 5. The coordinates of the hand are expressed by the following vector,

$$\vec{V}_{hand} = \begin{bmatrix} x_{hand} \\ y_{hand} \\ z_{hand} \end{bmatrix} \tag{1}$$

where x_{hand} , y_{hand} and z_{hand} correspond to the values of the coordinates in the three principal direction. Figure 5 represents also the thresholds that define the spatial limits of the neutral position. The designed gestures lie outside of these thresholds.

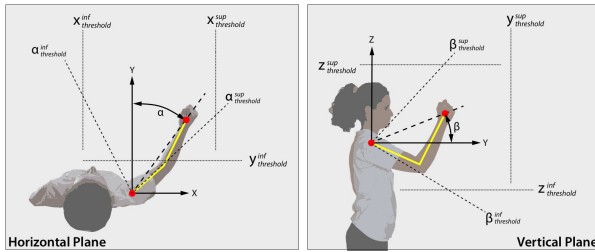


Fig. 5. Representation of the reference system used to detect the gestures. The yellow line indicates the user’s skeleton while the dashed lines are the thresholds used for the detection.

Since the gestures involve the motion of the hand all the time, its position in the space according to the defined coordinate system is used for the recognition. However, these data are not completely reliable due to different arm dimension and joints mobility among the users. The two angles α and β have been taken into account in order to overcome the issue related to the dimension. The angle α is defined as the rotation of the hand around the shoulder on the horizontal plane and it is possible to calculate it by the following equation:

$$\alpha = \pi/2 - atan2(x_{hand}, y_{hand}). \tag{2}$$

On the other hand, β is the rotation of the hand around the shoulder on the vertical plane. Its equation is :

$$\beta = \text{atan2}(z_{hand}, y_{hand}). \tag{3}$$

Once all of these measures are calculated, the algorithm detects if a user’s movement corresponds to one of the gestures according to the comparison represented in the scheme in Fig 6.

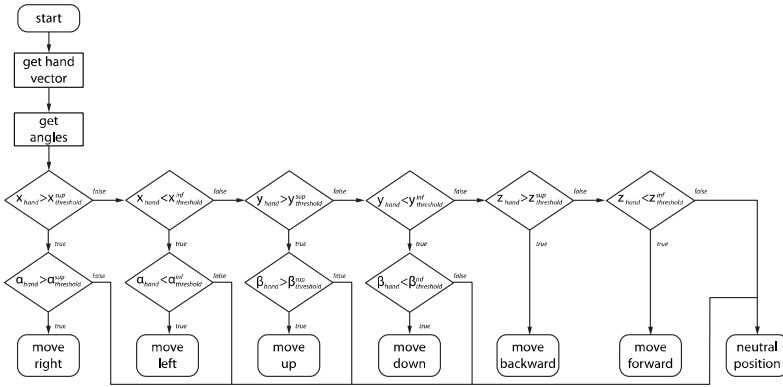


Fig. 6. The schematic representation of the detection algorithm

4.3 Gesture Definition

An experimental investigation has been carried out to define the values of the thresholds. Ten users (5 males, 5 females, height between 1.55 and 1.88m) were asked to perform the designed gesture in both of the directions. A Kinect was placed in front of the user and all the movements were recorded my means of FFAST. The measures were elaborated in order to find the extreme values reached by the hand and by the width of the angles. These data correspond to the end of the gestures, before returning to the neutral position. Each threshold corresponding to a gesture in a direction is related with these data. In particular, it is equal to the 85% of the minimum value between the peaks. Table 1 reports the values of the thresholds.

Table 1. The values of threshold

	Superior	Inferior
X	26.5 cm	-12.6 cm
Y	39.1 cm	11.0 cm
Z	28.5 cm	-32.4 cm
α	38.3 deg	-22.9 deg
β	83.6 deg	-45.5 deg

5 Case Study

This section reports a case study where the system is used in the field of furnishing. Nowadays, many e-commerce stores allow the customers to browse in large catalogues of objects, but they are limited on a planar navigation on a web-like application. Therefore, the searching activity does not differ much to the ones on a paper catalogue. The system developed makes the searching activity different to the traditional one because it is in a 3D space.

Figure 7 shows the forty-eight different objects that have been used to populate the environment. They are pieces of furniture addressed to housing and they are arranged in a virtual space according to three features. Tab. 2 gives a map of the objects. The first feature corresponds to the recommended location in the house where the object should be placed. In particular, the objects are addressed to three different places: kitchen, living room and bedroom. A vertical move of the camera in the virtual environment enables the user to switch between these three kinds of locations.

The second feature is the function of the objects. The environment is populated by furniture addressed to support the person, such as chairs and beds, to support other objects, like in case of tables, and to keep objects like a wardrobe. The user can navigate across the three functions by moving back and forward.

The last feature is the style used for designing the objects. The user can switch to different kinds of styles by moving left and right.

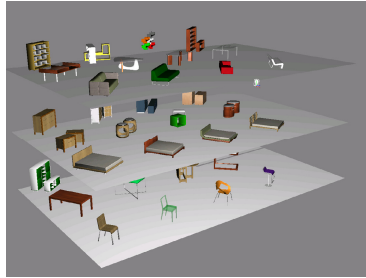


Fig. 7. The virtual catalogue

Table 2. The objects in the virtual space. For each object, models with different style have been arranged in the space.

		<i>Function</i>		
		Support people	Support objects	Keep objects
<i>Location</i>	Livingroom	Sofas	Coffe tables	Book cases
	Bedroom	Beds	Night tables	Wardrobe
	Kitchen	Chairs	Tables	Cupboard

As it is possible to see in Fig. 8, the user stands in front of the PowerWall and navigates the virtual catalogue by means of gestures. Once he finds the piece of furniture he wants, the user can select it by means of voice commands. In this case, a HUD appears in order to show to the screen auxiliary information, such as a brief description and also a list of further available colours (Fig. 8).

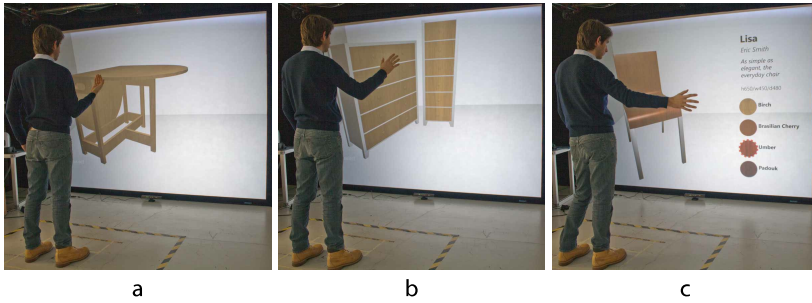


Fig. 8. The user during the case study

6 Discussion

The case study shows how the system can be used to navigate in a virtual environment with sorted objects by means of gestures. In particular, it reveals an approach to move across different pieces of furniture. The big advantages of this system are due to the integration of the 3D visualization and the natural interaction. The objects are organized in the space so that the user is able to search easily several of them. In a traditional desktop application for browsing information, the user sees a planar space and he can switch only from a selected object to the previous or the next one. In this case, instead, each virtual object is surrounded in all of the directions by other six. Therefore, the compact organization of the virtual items in the environment allows the user to move quickly his research towards what he is looking for. The gestures are simple to learn and it turns out the system is able to detects them precisely. The ease of gestures is given by the strict association between the user movement and the function. For these reasons, the developed system provides some benefits regards traditional approaches for browsing.

However, during the use of the system in the case study, a problem regarding the user orientation in the virtual space came out. The traditional interfaces give a clue to the user about his position between the items by means of visual feedbacks. The scrollbar usually provides this clue. The proposed system misses a graphic way to represent the user location in the 3D space and it informs the user only when he is at the last objects along a direction by means of a trembling movement.

7 Conclusion

The work in this paper describes the feasibility of a visualization and interaction system to navigate in a VR environment. The system is specifically addressed to environment in which the virtual objects are spatially organized on a three dimensional grid. The objects are arranged along the three spatial directions according to three different classifications.

The Authors believe this system can have a large use in many activities that involve searching information or items. Actually, the NUI technologies available today are spreading in many fields and can be integrated in several useful applications. In particular, the system can be easily integrated in commercial activities, in order to show products, as also demonstrated by the reported case study. The use of VR visualization and natural interactive gestures can increase the user involvement and consequently improve the user experience while seeking information.

Further studies can be carried out in this field. The Authors intend to continue investigating this kind of navigation. In particular, they want to develop new natural interaction techniques, by increasing and improving gestures. They want also to demonstrate the effectiveness of the system in other navigation contexts. Finally, comparative tests will be carried out in order to evaluate the user performances and satisfaction during the use of the system.

References

1. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: A survey. *Comput. Vis. Image Underst.* 108, 116–134 (2007)
2. Wachs, J.P., Kölsch, M., Stern, H., Edan, Y.: Vision-based hand-gesture applications. *Commun. ACM* 54, 60–71 (2011)
3. Cruz-neira, C., Sandin, D.J., Defanti, T.A.: Surround-Screen Projection-Based Virtual Reality The Design and Implementation of the CAVE. In: *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 135–142 (1993)
4. Bowman, D.A., McMahan, R.P., Ragan, E.D.: Questioning naturalism in 3D user interfaces. *Commun. ACM* 55, 78–88 (2012)
5. de Haan, G., Griffith, E.J., Post, F.H.: Using the Wii Balance Board™ as a Low-Cost VR Interaction Device. In: *Proceedings of the 2008 ACM Symposium on Virtual Reality Software and Technology*, pp. 289–290 (2008)
6. Moeslund, T.B., Hilton, A., Krüger, V.: A survey of advances in vision-based human motion capture and analysis. *Comput. Vis. Image Underst.* 104, 90–126 (2006)
7. Cabral, M.C., Morimoto, C.H., Zuffo, M.K.: On the usability of gesture interfaces in virtual reality environments. In: *Proceedings of the 2005 Latin American Conference on Human-Computer Interaction*, pp. 100–108 (2005)
8. Fang, Y., Chai, X., Xu, L., Wang, K.: Hand Tracking and Application in Map Navigation. In: *Proceedings of the First International Conference on Internet Multimedia Computing and Service*, pp. 196–200 (2009)
9. Suma, E.A., Lange, B., Rizzo, A.S., Krum, D.M., Bolas, M.: FFAST: The Flexible Action and Articulated Skeleton Toolkit. In: *IEEE Virtual Reality Conference*, pp. 247–248. IEEE (2011)
10. Roupé, M., Bosch-Sijtsema, P., Johansson, M.: Interactive navigation interface for Virtual Reality using the human body. *Comput. Environ. Urban Syst.* 43, 42–50 (2014)
11. Jeong, S., Song, T., Kwon, K., Jeon, J.W.: TV Remote Control Using Human Hand Motion Based on Optical Flow System. In: Murgante, B., Gervasi, O., Misra, S., Nedjah, N., Rocha, A.M.A.C., Taniar, D., Apduhan, B.O. (eds.) *ICCSA 2012, Part III. LNCS*, vol. 7335, pp. 311–323. Springer, Heidelberg (2012)

12. Wachs, J.P., Stern, H.I., Edan, Y., Gillam, M., Handler, J., Feied, C., Smith, M.: A Gesture-based Tool for Sterile Browsing of Radiology Images. *J. Am. Med. Informatics Assoc.* 15, 321–323 (2008)
13. Chai, X., Fang, Y., Wang, K.: Robust hand gesture analysis and application in gallery browsing. In: *IEEE International Conference on Multimedia and Expo*, pp. 938–941 (2009)
14. Ii, R.M.T., Hudson, T.C., Seeger, A., Weber, H., Juliano, J., Helser, A.T.: VRPNA Device-Independent, Network-Transparent VR Peripheral System. In: *Proceedings of the ACM Symposium on Virtual Reality Software and Technology*, pp. 55–61 (2001)
15. Nielsen, M., Störring, M., Moeslund, T.B., Granum, E.: A procedure for developing intuitive and ergonomic gesture interfaces for HCI. In: Camurri, A., Volpe, G. (eds.) *GW 2003. LNCS (LNAI)*, vol. 2915, pp. 409–420. Springer, Heidelberg (2004)