

# Developing and evaluating two gestural-based virtual environment navigation methods for large displays

Paulo Dias<sup>1,2</sup>, João Parracho<sup>1</sup>, João Cardoso<sup>1</sup>, Beatriz Quintino Ferreira<sup>2</sup>, Carlos Ferreira<sup>2,3</sup>, Beatriz Sousa Santos<sup>1,2</sup>

<sup>1</sup> DETI/UA- Department of Electronics, Telecommunications and Informatics

<sup>2</sup> IEETA- Institute of Electronics and Telematics Engineering of Aveiro

DEGEI/UA – Department of Economics, Management and Industrial Engineering  
University of Aveiro

Campus Universitário de Santiago, 3810-193 Aveiro, Portugal

paulo.dias@ua.pt; parracho@ua.pt; joaocardoso@ua.pt;  
mbeatriz@ua.pt; carlosf@ua.pt, bss@ua.pt

**Abstract.** In this paper we present two methods to navigate in virtual environments displayed in a large display using gestures detected by a depth sensor. We describe the rationale behind the development of these methods and a user study to compare their usability performed with the collaboration of 17 participants. The results suggest the users have a better performance and prefer one of them, while considering both as suitable and natural navigation methods.

**Keywords:** Navigation in Virtual Environments, Gestural interaction, 3DUIs.

## 1 Introduction and Motivation

The scope of 3D interaction has been expanding, creating new opportunities and challenges. One such opportunity is caused by the advent of large displays located in public spaces [1], which may be leveraged to interactively provide information or other functionality to persons walking by. In order to support students' assignments and to foster a better understanding of the issues involved in interaction with large displays we have been developing an interactive system, located at the entrance hall of our Department, including a large screen and a Kinect sensor, meant to run applications that might be used to display relevant information, making demos or just for fun [2]. Allowing a user to navigate through a virtual environment (VE) in a natural way that would let passing by users, for instance, easily take a campus virtual tour, was one of the main goals and thus an adequate navigation method was an important feature. Reviewing the related literature [3, 4], and exploring tools that might allow using Kinect as a 3D input device, two navigation methods were developed taking into consideration the application and context of use. The two methods were named and will be referred to as: "Bike" and "Free Hand". The rationale for these methods was the utilization of simple and natural gestures that neither involve very high concentration

nor effort of the user for the execution of the various actions, and are easy to learn. After an iterative process involving some formative evaluation carried out to improve the usability of the methods, a user study was performed to compare them.

The remaining of this paper is organized as follows: section 2 offers a summary of related work, section 3 presents the navigation methods, and section 4 describes the user study and presents the main results. Finally some conclusions are drawn in section 5.

## **2 Related work**

According to [1, 5] the 3D interaction methods go beyond the traditional/typical use in Virtual Reality; however, research of 3D UIs for non-VR environments is still in an early stage. Nonetheless, nowadays, 3D UIs seem to have found new opportunities in two different domains: gaming and public large displays. The latter are becoming larger, with higher resolution and with increased ubiquity [5], as well as more and more frequent, namely in public spaces; and if formerly displays showed information in a passive way, this paradigm is now changing and new user interfaces need to be designed for such context. Spatial input in 3D UIs enables users to interact with remote large displays freely, not needing any type of specialized input device or gear. Recent developments in computer vision have made it possible to detect free-hand gestures performed in the empty space using widely available and quite affordable hardware, such as the Microsoft Kinect. In fact, gestural methods to interact with large displays follow the novel trend towards “natural” user interfaces [5].

Previous works have already combined navigation and selection methods with spatial input in 3D UIs to interact with large displays [2, 3, 4, 6]. In the present work we focus on the development and evaluation of navigation methods considered as “natural” user interfaces.

Navigation in virtual environments usually is characterized by a user getting around within the environment [7] through the manipulation of a virtual camera and possibly an avatar to a desired position, simulating the humans’ movement in the real world and hence providing a feeling of immersion in the VE.

Regarding 3D UIs evaluation, formative and summative methods are widely used in different phases of the iterative development cycle [8], resorting to task performance as well as user satisfaction measures. To gather user satisfaction data, questionnaires and interviews are often used, whereas to obtain performance measurements, observation is most suited. Since gestural user interfaces are relatively recent and dissimilar from traditional 3D interfaces used in virtual reality systems, they pose specific issues during evaluation. In the addressed case of large displays, the specific issues are related to location, lighting conditions or other passing by users.

### 3 Proposed navigation methods

In order to allow users to navigate in a virtual environment in a natural way through gestures we developed two navigation methods dubbed “Bike” and “Free Hand”, both based on very simple metaphors [3].

The “Bike” method emerged as an evolution (based on a more common and realistic metaphor) of the method presented in [3] that proposed a “Broomstick” navigation. Indeed, our “Bike” method differs from the latter as the control for the direction is not related with the users’ shoulders but with the relative position of the hands.

On the other hand, the “Free Hand” arose from two practical motives. The first was to provide a sense of continuity and coherence relatively to the designed interface already in use for the rest of the application in the public large display (allowing namely to browse the faculty contacts list or access course schedules through movements of the dominant hand). Additionally, this “Free Hand” method offers a very similar interaction to the typical mouse-based interface, resulting in a familiar and easy user learning process.

#### 3.1 Bike

The “Bike” method uses a metaphor similar to the control of a bicycle, i.e., the user initiates the action by placing both hands alongside with closed fists as if to grab the handlebar of a bicycle (Fig. 1(left)). Thus, when the user puts their right hand slightly forward and the left hand back, the camera turns left. Changing the order of the hands, left hand in front and right hand back, it turns the camera right. The speed control of the forward (or backward) movement is done by advancing or pulling back both hands in parallel (Fig. 1(right)). To allow a larger range of speed, the user might also step forward or backward in order to get closer or further from the Kinect respectively increasing or decreasing the overall speed.

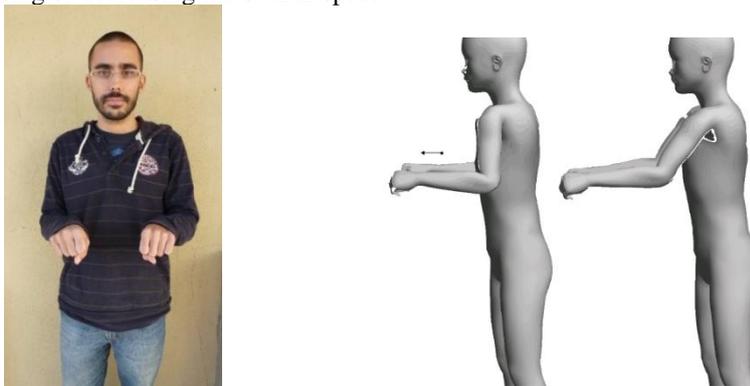


Fig. 1. Demonstration of “Bike” method gestures

### 3.2 Free hand

The “Free Hand” method was developed for consistency with the interaction methods used in other applications of our interactive system and is based on the fact that users usually control a cursor (mouse) with their predominant hand. The control of the view camera is done with the gestures of users’ dominant hand (Fig. 2). The navigation speed is controlled giving a step towards or away from the Kinect sensor; the bigger the step, the higher the speed of the movement.



**Fig. 2.** Demonstration of “Free Hand” method gesture

### 3.3 Development

Both navigation methods were developed using the Unity 3D platform. During the application loading the initial position of the hands of the user are set as reference for the following movements.

The technology used to implement the navigation methods was Unity and Microsoft Kinect SDK v1. In order to make both tools to communicate, a Unity package (Kinect wrapper for Unity) provided by the developer community was used.

In “Bike” method, the orientation between the hands as well as their relative position corresponding to the initial position is used to determine the movement (front/back – distance of hands to Kinect and left/right – right hand slightly in front of right hand or vice versa).

In “Free Hand”, the initial position is once more used as reference for the camera steady position. Movements (front/back and left/right) correspond to movements of the hand away from the reference/initial position in the same direction. The further from the reference position, the faster the movement in the given direction.

Unity is used to control the physics of the scene, namely to perform collision detection. Since the camera does not have physics intrinsically associated to it, a sphere is created around the camera position to allow collision detection between the camera and the scene.

## 4 User study

As a summative evaluation of the two proposed navigation methods, a user study was performed with the collaboration of 17 participants. In this section we briefly present the methods used as well as the main results.

### 4.1 Methods

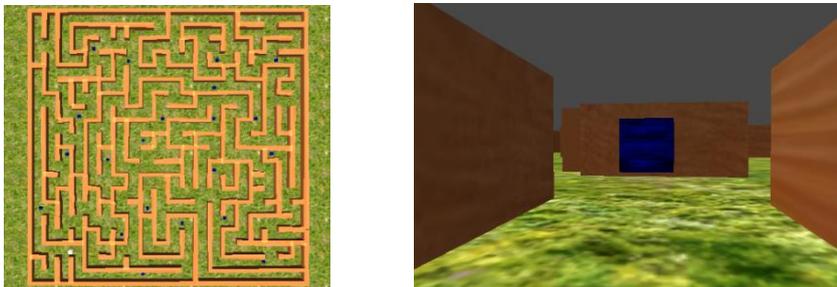
A simple maze was devised in order to test the performance of the users with both methods. Flying boxes were added to control progression within the maze and to give users a goal (catch the maximum number of boxes within the available time). Both the developed maze and the flying boxes are depicted in Fig. 3.

A within-subjects experimental design was used. The input variable for this experiment was the navigation method, with two levels, “Bike” and “Free Hand”. We assessed satisfaction through a post-task questionnaire, and performance based on the number of boxes caught, the number of collisions with the walls, and the velocity, as in previous similar user studies concerning navigation [9, 10]. The questionnaire addressed also specific aspects of the navigation methods, such as intuitiveness, need of training, and adequacy of control, including ten questions to be answered in a 5 level Likert-type scale, as well as the possibility to leave any comments or suggestions concerning the methods. Previous experience with Kinect or similar application was also registered.

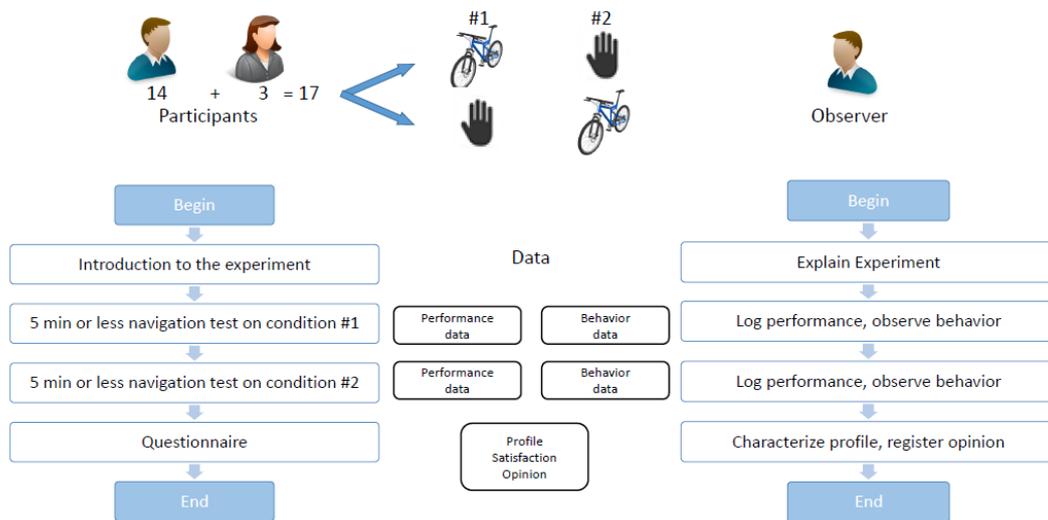
Sixteen students and a faculty member from our Department used the two methods to navigate the maze, for 5 minutes with each method, trying to capture as many boxes as possible. The experiment was performed in the lobby of our department where the system is running, and all participants were briefed concerning the two gestural methods and were allowed to train for a few minutes before the trial.

As a within-subject experimental design was used, we counterbalanced for possible effects of learning or boredom on the results, asking half of the users to start by one method and the other half by the other method. The followed protocol for the experiment is illustrated in Fig. 4.

Acquired data were analyzed using Exploratory Data Analysis, parametric tests, and non-parametric tests due to the relatively low number of the participants [11, 12].



**Fig. 3.** Aerial view of the maze and user view with a box to catch



**Fig. 4.** Experimental protocol: within-group design; input variable (with two levels): navigation method (“Free Hand” and “Bike”); output variables: performance and satisfaction; 17 participants.

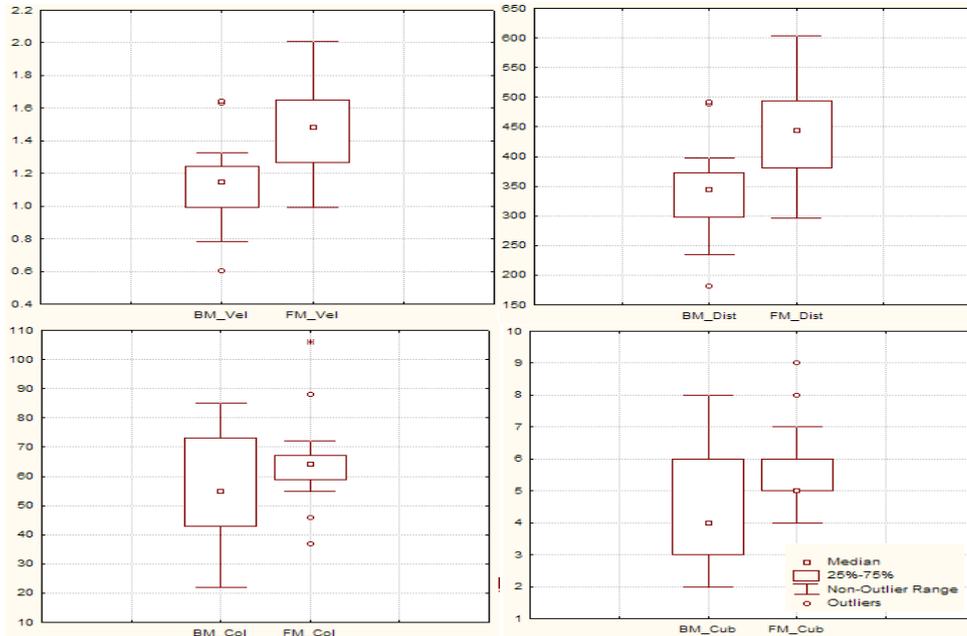
## 4.2 Results and discussion

As mentioned, seventeen users tested the system in a real setup, sixteen were aged 19 to 26 and one user was 38 years old, three participants were female and fourteen were male. All users were right handed.

Table 1 and Fig. 5 show the main results for the performance variables (measured in a ratio scale): speed, distance, number of collisions and number of caught boxes measured with the two navigation methods.

**Table 1.** Average and median of the logged performed variables

	<b>Bike</b>	<b>Freehand</b>
<b>Speed (average)</b>	1,13	1,49
<b>Distance (average)</b>	337,6	447,3
<b>Collisions (median)</b>	55	64
<b>Objects (median)</b>	4	5



**Fig. 5.** Boxplots of the logged performance variables (top left- speed; top right- distance; bottom left- number of collisions; bottom right- caught objects (cubes))

The median values of the number of caught objects were 4 with the “Bike” and 5 with the “Free Hand” method. A t-Student test as well as a Wilcoxon Matched Pairs test rejected the equality hypothesis (with respectively  $p=0.0102$  and  $p=0.0175$ ) meaning that the difference between the number of objects caught by the users with both methods is statistically significant and cannot be due to chance.

Also in the case of distance and speed, t-Student and Wilcoxon tests rejected the equality hypothesis (with  $p=0.0001$  and  $p=0.0008$  for distance; and  $p=0.0001$  and  $p=0.0008$  for speed). This might be justified by the following behavior observed throughout the experiment: most users in “Free Hand” just step forward and keep moving always at the same speed independently of the number of collisions. On the other hand, with “Bike”, most users stop the movement forward to perform the camera rotation resulting in a lower speed.

In contrast, the median values of the number of collisions (55 with the “Bike” and 64 with the “Free Hand” method) are not significantly different, since the above mentioned tests didn’t reject the equality hypothesis.

Based on these results we may conclude that users performed globally better when navigating with the “Free Hand” method as they caught more objects, attained higher speeds and traveled larger distances, with approximately the same number of collisions.

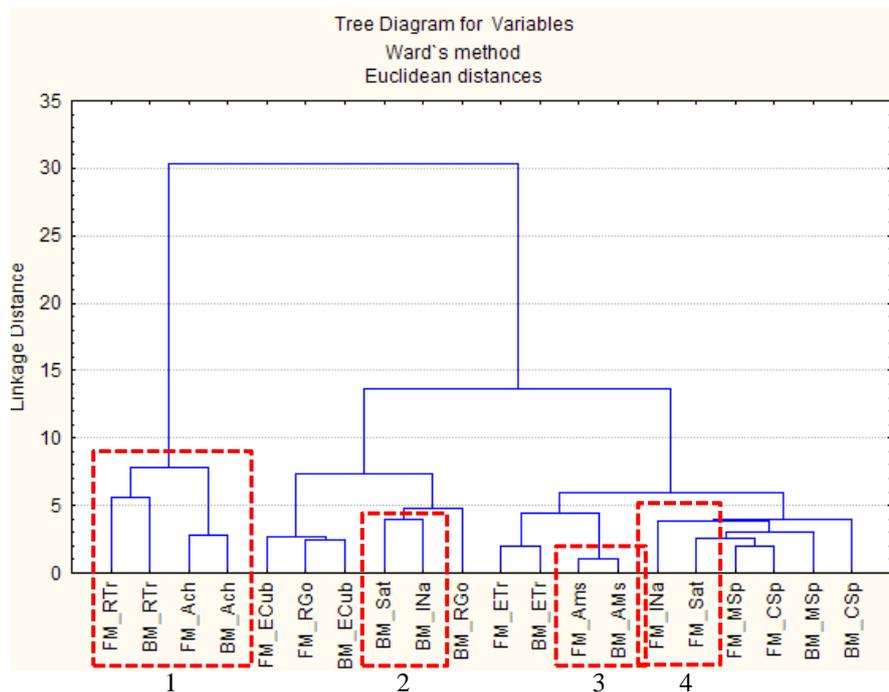
Fig. 6 depicts a dendrogram [13] representing similarity among answers to the questions concerning the two navigation methods. Box number 1 draws attention to the cluster of the variables “has annoying characteristics” (ACh) and “requires train-

ing” (RTr) which show a similar profile (low values) while more different from all the other variables in both methods, meaning that the former (ACh) might be an adequate proxy for the need of training (RTr). Moreover, their low values suggest that these aspects are considered suitable by users in both cases.

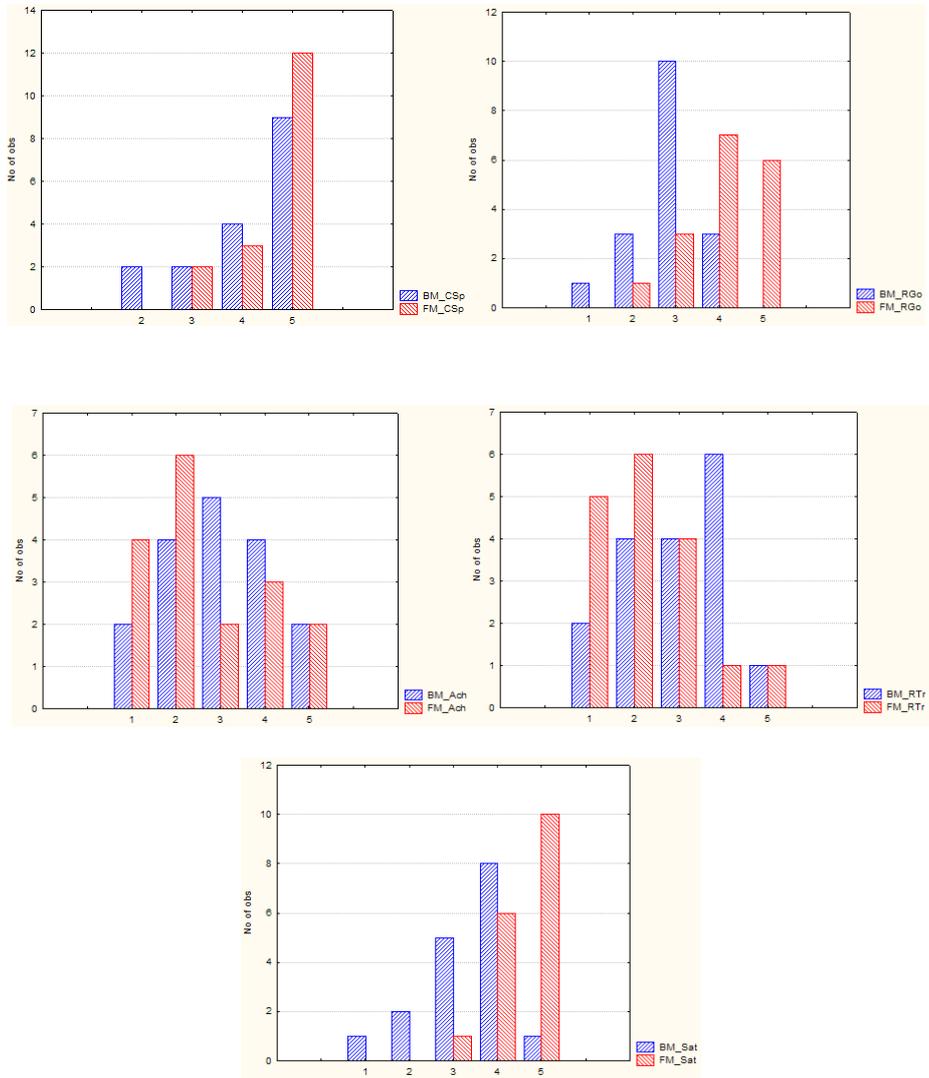
Boxes number 2 and 4 highlight the clusters formed by variables “intuitive navigation” (INa) and satisfaction (Sat) for “Bike” and “Free Hand” methods, respectively, suggesting a high correlation between the two variables, which might imply that intuitiveness is a fundamental characteristic of a navigation method.

Box number 3 points out that the users’ answers concerning application messages (variable AMs) were almost identical for both methods, meaning that there is virtually no difference between the feedback provided by the application in both cases.

Fig. 7 shows the main results of the post-experiment questionnaire concerning the two navigation methods: “Bike” (BM) (blue) and “Free Hand” (FM) (red). It depicts the bar charts of the users’ answers to the questions (in a 5 level Likert-type scale) that were significantly different for both methods, from left to right and from top to bottom: CSp - camera speed is adequate, RGo – easy to reach goal, ACh – has annoying characteristics, RTr - requires training, Sat - overall satisfaction. These ordinal variables were tested using Wilcoxon Matched Pairs test which rejected for all these five cases the equality hypothesis with the corresponding p values: CSp:  $p=0.0431$ ; RGo:  $p=0.0015$ ; ACh:  $p=0.0382$ ; RTr:  $p=0.0367$ ; Sat:  $p=0.0010$ .



**Fig. 6.** Dendrogram showing similarity among answers to the questions concerning the two navigation methods: “Bike” (BM) and “Free Hand” (FM). Boxes show noticeable cases



**Fig. 7.** Questionnaire results concerning the two navigation methods: “Bike” (BM) (blue) and “Free Hand” (FM) (red) for the questions that were significantly different for both methods (in a 5 level Like-type scale from 1- totally disagree to 5- totally agree) (from top-left to bottom-right CSp- camera speed is adequate, RGo- easy to reach goal, ACh- has annoying characteristics, RTr- requires training, Sat- overall satisfaction (1- not at all satisfied to 5 – very much satisfied))

## 5 Conclusions

In this work we address the development and evaluation processes of two gesture-based virtual environment navigation methods designed for interaction with a large display.

Throughout the experiment, a similar interest in both methods was noticed by the experimenter. While the users' performance and satisfaction were significantly better in some of the measured variables with the "Free Hand", we believe that users also considered the "Bike" method as suitable and natural for navigation, and in retrospective we understood that the main constraint of the latter was that users could not stop the interaction efficiently. This is in line with the "non-parkable" issue pointed in [5], which hampers increasing precision in spatial/free-hand 3D interfaces.

Fatigue was not considered in this work given the limited duration of interaction in our application. However for longer interactions this factor should be considered since more tiring gestures might be less adequate.

The affordance provided by the used metaphor, a bicycle handlebar, may be explored visually and is an envisaged future work direction. Such discoverability of possible actions is of the utmost importance since these methods are to be implemented on public displays applications, requiring a self-explanatory user interface, where the visual representation of a bicycle handlebar or a steering wheel or even an avatar of the users' hands, may indicate passing-by users the initial form of interaction.

**Acknowledgements.** The authors are grateful to all volunteer participants. This work was partially founded by National Funds through FCT - Foundation for Science and Technology, in the context of the project PEst-OE/EEI/UI0127/2014, and by the Program "Partilha e Divulgação de Experiências em Inovação Didática no Ensino Superior Português" (84/ID/2014).

## References

1. D. A. Bowman, S. Coquillart, B. Froehlich, and M. Hirose, "3D User Interfaces : New Directions and Perspectives," *IEEE Computer Graphics and Applications*, vol. 28, no. 6, pp. 20–36, 2008.
2. P. Dias, T. Sousa, J. Parracho, I. Cardoso, A. Monteiro, and B. Sousa Santos, "Student Projects Involving Novel Interaction with Large Displays," *IEEE Computer Graphics and Applications*, vol. 34, no. 2, pp. 80–86, 2014.
3. G. Ren, C. Li, E. O'Neill, and P. Willis, "3D freehand gestural navigation for interactive public displays," *IEEE Computer Graphics and Applications*, vol. 33, no. 2, pp. 47–55, 2013.
4. M. N. Kamel Boulos, B. J. Blanchard, C. Walker, J. Montero, A. Tripathy, and R. Gutierrez-Osuna, "Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation," *International Journal of Health Geographics*, vol. 10, no. 45, 2011.

5. D. A. Bowman, "3D User Interfaces," in *The Encyclopedia of Human-Computer Interaction*, M. Soegaard, and Rikke Friis Dam, 2nd ed., Aarhus, Denmark: The Interaction Design Foundation, 2014, chapter. 32. [https://www.interaction-design.org/encyclopedia/3d\\_user\\_interfaces.html](https://www.interaction-design.org/encyclopedia/3d_user_interfaces.html)
6. G. Ren, and E. O'Neill, "3D selection with freehand gesture," *Computers and Graphics*, vol. 37, no. 3, pp. 101-120, 2013.
7. J. Jankowski and M. Hachet, "Advances in Interaction with 3D Environments," to appear in *Computer Graphics Forum*.
8. D. A. Bowman, E. Kruijff, I. Poupyrev, and J. LaViola, *3D User Interfaces: Theory and Practice*, Addison Wesley, 2005.
9. B. Sousa Santos, P. Dias, A. Pimentel, J. W. Baggerman, C. Ferreira, S. Silva, and J. Madeira, "Head-mounted display versus desktop for 3D navigation in virtual reality: a user study," *Multimedia Tools and Applications*, vol. 41, no. 1, pp. 161–181, Aug. 2009.
10. J. Lapointe, P. Savard, and N. G. Vinson, "A comparative study of four input devices for desktop virtual walkthroughs," *Computers and Human Behaviour*, vol. 27, no. 6, pp. 2186–2191, 2011.
11. D. Hoaglin, F. Mosteller, and J. Tukey, *Understanding robust and exploratory data analysis*. New York: Wiley, 1983.
12. T. Hettmansperger and J. McKean, *Robust nonparametric statistical methods* (Kendall's library of statistics. vol. 5). London, U. K.: Arnold, 1998.
13. J. Hair Jr., W. C. Black, B. J. Babin, and R. E. Anderson, *Multivariate Data Analysis – A global Perspective*, 7th ed. Upper Saddle River: Pearson Education, 2008.