

A comparative study of walkthrough paradigms for virtual environments using Kinect based natural interaction

Luis A. Hernández-Ibáñez, Viviana Barneche-Naya, Rocío Mihura-López
videaLAB
Universidade da Coruña
A Coruña, Spain
{luis.hernandez; viviana.barneche; rocio.mihura}@udc.es

Abstract— This study analyzes and compares several gestural schemes using Kinect based interaction on virtual walkthroughs. The text describes the tests performed to measure the suitability of six different combinations of body gestures for moving and turning, and compares the results in terms of usability. The tests were carried out on a real existing virtual museum installation.

Keywords— natural interaction; Kinect; virtual walkthrough; virtual museum

I. INTRODUCTION

The problem of controlling the movement of a user inside a virtual environment without using apparent physical interfaces exists since the arising of Virtual Reality installations in the 90's. Authors like Bowman [1] have already pointed out that the detection and interpretation of user's natural gestures could be used for implementing forms of movement inside virtual models. The capability of those VR installations to grasp user's gestures was very limited in those times considering that only gaze direction (obtained from HMD orientation) as well as hand location and orientation (taken by special pointing devices) could be used to interpret user's desires.

Nowadays, depth camera technologies, such as the Kinect® system, facilitate in obtaining a very comprehensive description of user's pose and gestures. Therefore, this device is being used profusely in multifarious fields, where virtual environments have to be explored by means of natural interaction. Among those fields, Architectural Visualization and Virtual Museums demand good walkthrough paradigms for exploring the space and contemplating the environment and the objects on display, prior to enabling further interactions [2].

There are several interesting examples developed in recent years, which implement different walkthrough paradigms [3][4][5][6][7][8]. Some of them support their validity offering tests of usability and results of their UX analysis, but there exist important conceptual differences among them. Until now, one cannot find a real study that contrasts the results of the use of several of them applied to the same general case. Such study could throw light on their suitability

in relation to the criteria utilized on interaction design for every virtual environment.

Virtual walkthrough requires the combination of two different groups of gestures. The first of them is related to movement along the scenario, changing the representation of the user's position in the virtual world by moving forward and, in some cases, allowing moving backward or sideways as well. The other group of gestures deals with changing user's orientation in the virtual environment, affecting the user's direction of displacement, which is mostly coupled with the view direction. Other gestures can be used for further interaction like selecting, manipulating objects, etc.

The catalogue of possible gestures is extensive. Apart from the cases found in existing examples, other possible gesture combinations have been proposed [9]. This study measures and compares six different gestural walkthrough schemes. Five of them can be found in previous works developed by different teams, while the authors of this paper developed the sixth scheme to accomplish this work in a befitting manner.

Those six walkthrough schemes combine a user's gesture to move forward with another gesture of turning. Based on the aforementioned cases, the gestures analyzed included point with arm for moving and turning [3][4], step forward/point with arm [3], right arm forward/body rotate [5], swing arms/body twist [6], lean forward/body twist [7], and step forward/body twist[8].

II. METHODOLOGY

All movement schemes were implemented on a common virtual environment depicting a reconstruction of a 4th century Roman villa, which is actually used as a virtual museum in a real world installation. Such virtual architecture provided a good and comprehensive set of scenarios, which were used for obtaining all the measurements required for the usability tests, being at the same time an entertaining and educative experience for the users involved, hence also allowing the authors to obtain results about the suitability of every paradigm for this kind of virtual museums.

The virtual villa was implemented in the popular game engine *Unreal Engine 4*. The Kinect based interaction was coded inside the system using the K4U libraries developed by Opaque Multimedia. The test set consisted of a low lit room with a projection screen having a Kinect sensor underneath, and some marks on the floor; one of them indicating the starting point of the experience, located 3.80 meters in front of the screen.

A. Test subjects

A group of 28 people took part in the experiment, having age ranging from 18 to 52, in equal proportions (50%, 50%) of female and male subjects. They were asked about their previous experience using videogames (39.3% were frequent users, 42.9% rarely played; 17.9% never played them).

The subjects tried different movement schemes in random order. They bestowed their subjective opinion using the provided questionnaires, and all additional comments were also recorded. The system also monitored and recorded their moves to extract additional data that would be contrasted against their opinion.

B. Test mechanics

The test for each movement scheme consisted of five stages:

1) *Training path*: The test began with an initial easy walkthrough crossing the villa. The user followed a path indicated with direction arrows in the turning points. This path included entering the villa through its main door, (Fig.1) crossing the vestibule to the main atrium, surround it and exiting the atrium through a doorway located in the side opposite to the entrance until they reach another exit leading to the garden. To accomplish this, users had to align with doors and doorways and make several 90° turns in both the directions. The mission of this first stroll was to facilitate in evaluating the ease of learning by measuring the time used, the number of gestures needed to accomplish the walk and counting the collisions detected and the time that the user remained in a collision state (i.e. sliding against a wall).



Fig. 1. View along the training path. Notice the arrow sign on the column.

2) *Free walk*: Afterwards, the user took a two minutes free walk to contemplate the house and the elements exhibited inside.

3) *First questionnaire*: The user answered a short questionnaire related to his or her impressions using the movement scheme used. Users could provide scores from 1 to 10 to every aspect related to ease of use, amount of attention put on controlling the movement, comfort, response to the user's intention, fatigue, etc.

4) *Precision test*: Next, the user performed another walkthrough to help measure the precision of every paradigm. This one required a more precise driving, since the user was asked to pass between two rows of objects, turn around a well and pass between another two rows of objects oriented in a different direction to finish stopping in a given place. (Fig.2)



Fig. 2. View of the precision test area.

Again, we measured the number of movements, both accelerations and turns, the time used to complete the task, the number of collisions detected and the number of frames in a collision state. This way, we could contrast the user's subjective opinion against the measures taken of his or her movement behaviour.

5) *Second questionnaire*: Finally, the user answered some final questions regarding general aspects of the experience.

C. Gestures

The user gestures implemented in the system can be divided into two groups; the first group, which we may call *march gestures* related to start, maintain and stop the walk, with constant or varying speed, and the second group, which we may call *turn gestures* related to changing the walking direction. Both sets of gestures were combined to configure different schemes involved in the study.

1) *March gestures*:

a) *Point forward with arm*: The user is told to raise his or her hand to start moving and release it to stop. The system measures the angle formed by the elbow-wrist vector with the vertical. This angle must be greater than a given threshold in order to start moving. This threshold is established to avoid unintentional starts due to the fact that this angle is greater than zero in the idle pose of many people. As the user rises his or her hand, the speed increases until it reaches the maximum walk speed for the test at an angle of 45 degrees.

b) **Lean forward:** The user is instructed to lean forward slightly to begin the walk, now at the max speed, and straighten up to stop. The system analyses the vector with origin in the base of the neck pointing forward and compares it to the horizontal. Again, a postural threshold value is considered, and the walk begins once the user leans past this threshold.

c) **Swing arms:** The user has to swing his or her arms back and forth like walking. The movement starts immediately at a constant, maximum walking speed. The speed is decoupled to the frequency of swinging, so its value remains same independent of the differences in the user's walking paces.

d) **Step forward:** There are several marks on the floor; one straight line, parallel to the screen that indicates the starting position, and some arrow shaped marks pointing forwards and backward. The user is told to do one step ahead to begin marching, go one step back to the starting position to stop and even do another step back to move backwards. The speed varies depending on the distance to the starting position, hence it is possible to move faster by stepping forwards twice. Nevertheless, we configured the system to reach the maximum speed with a single, regular step forward. Once the march begins, it will not stop until the user steps back to the starting position. Thus, in order to facilitate the changes in the direction, the speed is reduced when the system detects a turn gesture, proportionally with the magnitude of the turn.

2) Turn gestures

Two different gestures were implemented to control the view and moving direction, which are always coupled in this installation.

a) **Point sideways with arm:** The user points his or her hand left or right for steering. The deviation from the initial idle angle determines the magnitude of the turn in radians per second up to a maximum angular speed. The arm's sideways movement must go past a threshold angle to start the turning mode.

b) **Twist upper body:** The user has to twist the upper body following the natural rotation that occurs while changing the walking direction. The system measures the angle between the vector that connects both shoulders and the screen plane and applies an angular speed to the user's camera based on that angle. Again, there is a threshold angle and a maximum angular speed.

The magnitude of the effect of every gesture into the user's movement depended on several coefficients (gesture amplitude, threshold angles, minimum displacements, etc.) Those coefficients were determined in previous usability tests.

3) Movement schemes

The experiment implements the six aforementioned movement schemes by combining a marching gesture and a turning gesture. Although all of them are used in the listed examples, yet there are some differences that have to be noted. For instance, the system described in [4] is part of an immersive VR environment, so the user may point in any

direction, hence establishing instantly the absolute moving direction. In our case, with the view limited to the screen space, the arm direction in the point/point scheme indicated relative turning direction.

Table 1 summarizes the combinations and displays the naming convention used in the graphs in this paper.

TABLE I. MOVEMENT SCHEMES

Action	Scheme name					
	1 PP	2PT	3LT	4SP	5WT	6ST
March	Point	Point	Lean	Step	sWing	Step
Turn	Point	Twist	Twist	Point	Twist	Twist

III. RESULTS

All movement schemes proved to be valid for controlling a virtual walkthrough, especially for paths not too complex and for short periods of virtual exploration. Nevertheless, differences exist and can be relevant depending on the features and requirements of the virtual experience. Some schemes would be more adequate for longer experiences, and some are more precise than others as we will discuss later.

A. Ease of learning

The following graphs illustrate the response of the users to every movement scheme. The designers verified that the training path could be completed using only 10 movements, including accelerations and turns. Since it was the first time that every user tried every scheme, the number of movements utilized could be a clue of the ease of learning of the system. The LT scheme obtained the best results in this field with a mean value of 15 movements. On the other side, WT and ST mean values doubled this number. (Fig 3)

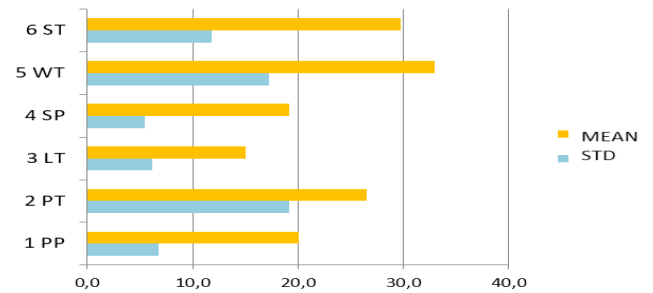


Fig. 3. Number of movements used to complete the training path

The analysis of the number of collisions and time in collision state helps to figure out of how quickly the user learns to drive the system properly (Fig.4). Fewer collisions indicate a good skill. High number of frames in collision state indicates that either the user advances stuck to the walls instead of returning to the center of the path or the user remains stuck against obstacles for a while.

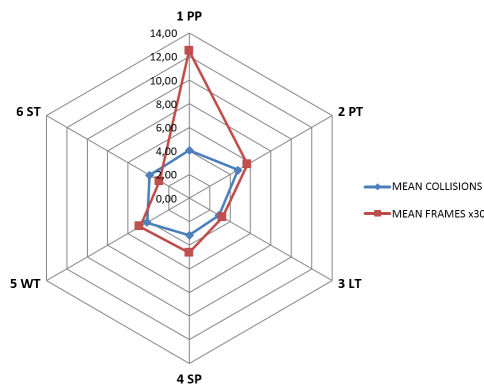


Fig. 4. Number of collisions and time in collision state along the training path

B. Ease of use

The results from the users' answers to the question regarding the ease of use of the system provided the highest scores (8.5) for the LT and SP schemes, and the worse qualifications for PP and WT. The LT scheme also obtained the best results regarding time used, number of movements and number of collisions. (Fig.5)

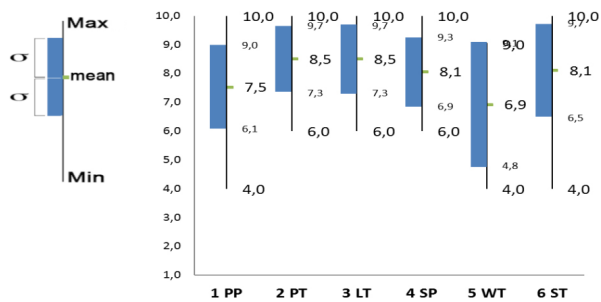


Fig. 5. Users' subjective response to the "ease of use" question

C. Attention

One of the goals of natural interface design is to develop systems that interfere as less as possible with the user's experience, responding to user's desires in a fluent, comfortable and confident way, allowing the user to focus attention on the experience instead of on the interface.

We asked the users regarding the amount of attention that they had to invest to drive the walkthrough instead of contemplating the environment. They were suggested to express it as a ratio of their total attention. (Fig. 6) describes the results.

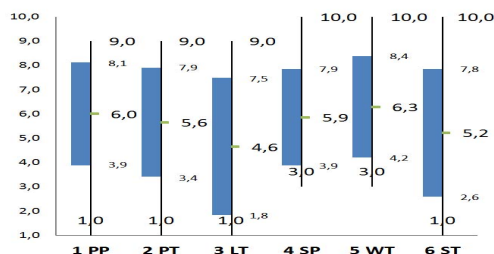


Fig. 6. Users' response to the "ratio of attention paid to system" question

Only the LT scheme obtained a mean value under 0.5, although it has the greater standard deviation. The WT scheme got the worst results in this question.

D. Fatigue

The question regarding the feeling of tiredness after using the system provided the results depicted in figure 7. All schemes obtained low values even lower for those that required less body and limb movements, such as the ST scheme.

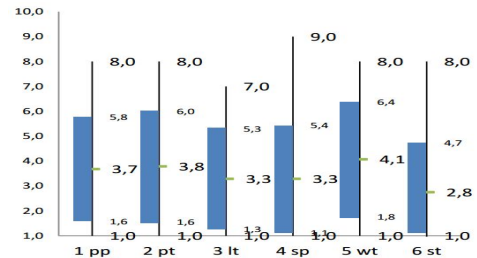


Fig. 7. Users' response to the question regarding fatigue.

E. System response to user's intentions

The users were asked about how well the system interpreted their intentions. The mean values of the responses are high in general. The LT scheme stands out again. (Fig. 8)

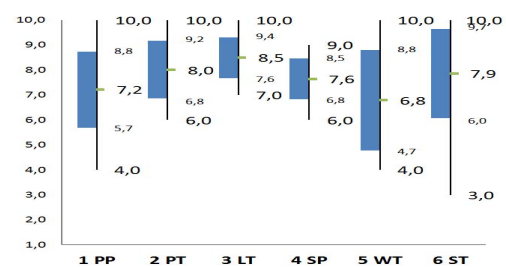


Fig. 8. Users' perception of system performance interpreting their intentions

F. Comfort

The next question was asked about how comfortable they felt using the system. Comfort perception relates (but is not limited to) to the user's sensations of ease of use and the effort put to control the movements. (Fig. 9)

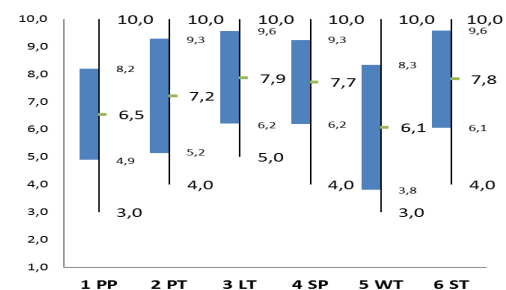


Fig. 9. System comfort for every scheme according users' perception

Specifically, we asked for their general sensation on this aspect after the test. The results assign the best values to LT and ST, being the worse for WT and PP, which make a more intense use of the arms.

G. Precision

As aforementioned, the second part of the test was mainly designed to check the performance of every movement scheme in terms of precision. The measures taken for this test are shown in the following figures. They include time taken to complete the path (Fig.10), number of movements (Fig 11), number of collisions and time in collision state. (Fig.12)

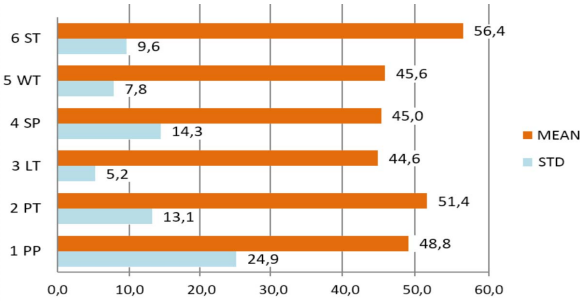


Fig. 10. Time in seconds used to complete the precision test.

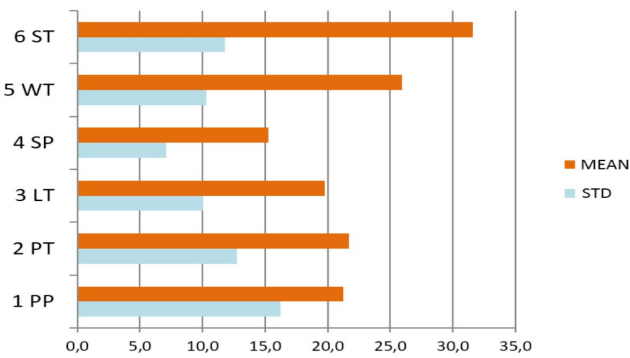


Fig. 11. Number of movements used to complete the precision test.

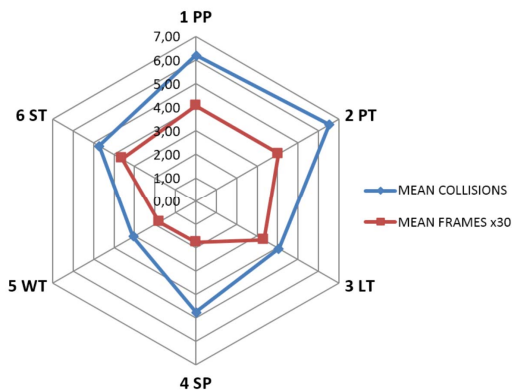


Fig. 12. Collisions and time in collision state during the precision test.

The previous graphs indicate that WT, LT and SP obtained the best results as the fastest schemes to complete the precision path, WT also permitted to do it with less collisions and smaller collision time, but the opinion of the users (Fig. 13)

assign the worst value to this scheme, probably due to the fact that it requires more movements, and give the best qualification to SP.

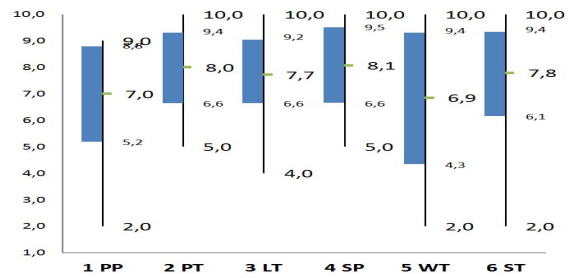


Fig. 13. Users' response to the question regarding precision.

H. General satisfaction

Upon finishing the test, users answered the question related to their general satisfaction with the usage and performance of the corresponding movement scheme (Fig.14). Four out of the six schemes got scores within a range of 0.1 points, so we can consider that PT, LT, SP and ST obtained the similar best results. PP and specially WT obtained the worst scores.



Fig. 14. Users' response to the question regarding general satisfaction.

IV. DISCUSSION

The analysis of the previous results, supported with the users' comments and the notes taken by the authors during the experiment, provides clues to characterize the behavior and performance of the movement schemes and their suitability for their use on a museum environment for virtual walkthroughs.

A. Point forward/Point sideways:

Although this is apparently the most clear and easy to use scheme, yet the results indicate a poor performance in general and a bad response from the users in terms of comfort, satisfaction and almost every other aspect.

The authors found other problems related to this scheme discussed in the preceding paragraphs:

- Different users interpreted the instructions in different, sometimes unexpected, ways. The apparently simple verbal instructions were: “Raise your arm to move forward. Point sideways to turn”. Some users raised their arms a lot, while others only a few. Some users raised only the forearm forming a right angle with the arm that stayed vertical. Some of them expected the system to respond to small movements, others moved their arms exaggeratedly. Consequently, the response of the system in terms of speed and acceleration was also different.

Considering that the instructions to use a natural interface should be extremely short and plain (if there should be any instructions at all), the fact that the ideal angles for the operation of this scheme were out of a short description, and the lack of unique interpretation were clearly a burden for the application of this scheme in a museum environment.

- Other problem arose when the user pointed in a direction coincident to the position of the Kinect device. The depth camera lost the track of the articulations, since the hand obscured the elbow in that gesture configuration. This happened several times during the experiment. Thus, if this scheme is to be used, we encourage placing the Kinect device over the screen instead of below it.

B. *Point forward / Twist upper body*

This scheme provided medium and low scores in almost all aspects, with the exception of the subjective responses to the questions related to ease of use, precision and general satisfaction. In this respect, the users’ answers to these questions do not correlate with the objective results. This seems to indicate that the PT scheme is perceived by the users as being more precise than it really is, probably due to the fact that they were not informed of the collisions that they incurred.

In a museum environment, this scheme could have a good general acceptance, but results suggest that it should be used in virtual environments where collisions do not represent a problem.

C. *Lean forward / Twist upper body*

This scheme scored the best results in 8 out of the 12 evaluated aspects and obtained medium qualifications only in the precision part. The standard deviations of the users’ responses were also particularly smaller than the others. Users considered it to be one of the easiest to learn and use, and more comfortable. It also obtained outstanding results in the aspects of good response to users’ intentions and less attention required. Thus, we can consider it to be the less intrusive of all.

The reason why this system produces a more uniform response can be related to the fact that the gestures utilized to control the movement are similar in magnitude for all users, thus generating very similar responses.

Some users explicitly manifested a higher sense of immersion while using this scheme. Globally, this is the scheme that gathered the best results.

D. *Step forward / Point with arm.*

This can be considered a medium class scheme. It obtained proper results in all aspects, with nothing especially outstanding but also without any clear flaw.

Like the other step based schemes, it has slight disadvantage of needing some sort of mark on the floor to indicate the starting position.

E. *Swing arms / Twist upper body*

This is the scheme that obtained the worst qualifications as a whole. It obtained the worst scores in 6 out of 12 measures.

Although several users found it very funny at the beginning, soon it proved to produce unnecessary tiredness.

It is worth noting that this system obtained good results in terms of collision, collision times and time to complete the precision test, thus suggesting that it is a precise scheme, but the amount of movements needed to obtain such precision and the fatigue that it generates make the users evaluate it as the less precise of all.

F. *Step forward / Twist upper body*

This scheme constitutes a good candidate for some kind of installations. It is a medium to high valued case in all fields, but it stands out especially in the aspects of comfort and less fatigue. It also yields good results in the questions regarding response to users’ intentions and overall satisfaction.

These scores make this scheme a good candidate for virtual environments that require medium to long times for exploration.

V. CONCLUSIONS

All the schemes tested in this experiment showed pros and cons and yielded different outcomes in terms of ease of learning and use, intrusiveness, interpretation of users’ intentions, comfort, and precision. The analysis of the results can help in clarifying the adequateness of every system for every particular museum installation based on its special features.

As a future line of research, the results obtained may be put more closely in relation to four aspects of UX; utility, learn-ability, efficiency and stimulation [10]. The authors expect that the performance metrics and UX results presented here would be useful for designers of virtual environments in order to choose the natural interaction walkthrough scheme that could best fit their needs basing on the particular features of their installation.

REFERENCES

- [1] D. A. Bowman, D. Koller and L. F. Hodges, «A methodology for the evaluation of travel techniques for immersive virtual environments,» *Virtual reality*, vol. 3, n° 2, pp. 120-131, 1998.
- [2] D. A. Bowman, D. B. Johnson and L. Hodges, «Testbed evaluation of virtual environment interaction techniques,» *Presence*, vol. 10, n° 1, pp. 75-95, 2001.
- [3] L. D. Souza, I. Pathirana, D. Mcmeel and R. Amor, «Kinect to Architecture,» *IVCNZ*, Auckland, New Zealand, 2011.
- [4] F. Sen, L. Diaz and T. Horttana, «A novel gestur-based interface for VR simulation: re-discovering Vrouw Maria,» in *Virtual Systems and Multimedia*, Milan, 2012.
- [5] H. Richards-Rissetto, J. Robertsson, J. Von Schewerin, G. Agugiaro, F. Remondino and G. Girardi, «Geospatial virtual heritage: a gestured based 3D GIS to engage the public with Ancient Maya Archeology,» *Archeology in the digital era*, pp. 118-130, 2014.
- [6] M. N. K. Boulos, B. J. Bkanchard, C. Walker, J. Montero, A. Tripathy and R. Gutiérrez-Osuna, «Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation,» *International Journal of health geographics*, vol. 10, n° 1, p. 1, 2011.
- [7] M. Roupé, P. Bosch-Sijtsema and M. Johansson, «Interactive navigation interface for virtual reality using the human body,» *Computers, Environment and Urban Systems*, vol. 43, pp. 42-50, 2014.
- [8] L. Hernandez-Ibáñez, V. Barneche-Naya and R. Mihura-López «Natural interaction and movement paradigms. A comparison of usability for a kinect enabled museum installation» *Lecture Notes in Computer Science*, vol 9753, pp 145-155, 2016
- [9] P. Dam, M. Loaiza, L. Soares and A. B. Raposo, «A study of selection and navigation techniques using Kinect in VR,» *Proceedings of the 14 symposium on virtual and augmented reality*, Niteroi, Brasil, 2012.
- [10] M. Thüring and S. Mahlke, «Usability, aesthetics, and emotion in human-technology interaction.» *International Journal of Psychology*. vol. 42, pp. 253-264. Psychology Press, London (2007)