Comparative Evaluation among Diverse Interaction Techniques in Three Dimensional Environments

Giannis Drossis[1,2], Dimitris Grammenos[1], Maria Bouhli[1], Ilia Adami[1], and Constantine Stephanidis[1,2]

[1] Foundation for Research and Technology - Hellas (FORTH), Institute of Computer Science N. Plastira 100, Vassilika Vouton, GR-70013, Heraklion, Crete, Greece
[2] University of Crete, Computer Science Department, Heraklion, Crete, Greece
{drossis,gramenos,bouhli,iadami,cs}@ics.forth.gr

Abstract. This paper reports on the results of a user-based evaluation that was conducted on a 3D virtual environment that supports diverse interaction techniques. More specifically, the interaction techniques that were evaluated were touch, gestures (hands and legs) and the use of a smart object. The goal of the experiment was to assess the effectiveness of each interaction modes as a means for the user to complete common tasks within the application. A comparison is attempted in order to provide an insight to the suitability of each technique and direct future research in the area.

Keywords: multimodal interaction, 3D user interfaces, gestural interaction, usability evaluation, comparative evaluation.

1 Introduction

In recent years a lot of scientific effort has been placed in the development of multimodal interaction techniques which allow users to interact with systems in non-traditional ways. Such techniques have shown promise in enhancing the user experience by allowing more natural interaction between the user and the system. The different interaction approaches often seen, apply dissimilar practices including wearable equipment such as head mounted displays, non-instrumented user tracking using cameras, tangible artifacts and desktop-based interaction.

This paper aims to assess and compare diverse modes of interaction in the demanding area of 3D environments [2], where the six degrees of freedom constitute an additional impediment as it requires extended interaction vocabulary, in order to provide an insight on the pros and cons of each approach. The means of interaction to be evaluated were selected bearing in mind the extent to which they are both affordable and natural to the users. In this direction, desktop interaction using a touch screen was chosen as a widely adopted and intuitive solution, enhanced with a tangible object (SmartBox) that complements navigation in 3D spaces. Furthermore, gestural and kinesthetic interaction is applied as a more natural expression beyond the limits of computer systems.

N. Streitz and C. Stephanidis (Eds.): DAPI/HCII 2013, LNCS 8028, pp. 3–12, 2013.

2 Background and Related Work

Jaimes et al. [6] define a multimodal system as a system that "responds to inputs in more than one modality or communication channel, such as speech, gesture, writing and others". Multimodal user interfaces support interaction techniques which may be used sequentially or concurrently, and independently or combined synergistically. According to Oviatt [11], "Multimodal interfaces process two or more combined user input modes (such as speech, pen, touch, manual gesture, gaze, and head and body movements) in a coordinated manner with multimedia system output".

Gesturing is a common approach which has been proven to be very intuitive to users and is widely used in literature [5, 8 and 17]. Gestures can be defined as a form of non-verbal communication in which visible body actions communicate particular messages. Hand gestures can be used to augment systems and allow additional interactions when combined with other means of interaction such as simple touch [14]. Gestures may not be limited to multi-touch and hands, but may be applied to feet as well: foot-based gestures are proposed by [1, 5 and 12] as an alternative interaction mechanism. Valkov et al. [16] use foot gestures to expand simple multi-touch interaction and boost navigation in dynamic and complex 3D Virtual Environments.

Body movement indicates the pose of a user's body as mentioned by Jaimes et al. [6], which can be tracked and applied for selective interaction with the environment, where the system may interpret a specific body pose in order to enable interaction in a specific manner. Papadopoulos et al. [13] use defined body poses recognition in order to allow navigation in 3d environments. According to their approach, whenever a user poses in a certain way, manipulation of a camera in a virtual 3d environment begins. Grammenos et al. [4] use the users' positions to visualize information according to the location of any user inside a room: each user can choose a topic by moving to the side and explore different information details by moving forward and backwards.

Finally, smart artifacts deliver natural, tangible interaction using predefined actions (e.g., pressure) through their embedded technology. For instance, accelerometers and magnetometers are used by [7, 15] in order to detect the orientation of items.

3 System Overview

3.1 System Design

The system the users assessed involves 3D interactive information visualization in the form of a timeline presented in two distinct views. The first view involved information visualization in the form of events placed on a two dimensional plane with time (expressed in periods) extending on the horizontal axis [Fig. 1, left]. The second view presented the same information complementarily, using the metaphor of a time tunnel, where time extends along the tunnel length. The events included multimedia information including text, images, videos and 3D models. Furthermore, the system provided the separation of events in categories as a filtration mechanism. The individual components of the timeline (such as its title and the available categories) held the same appearance in both views for consistency.

The first view of the timeline aimed to provide an overview of the events displayed and offer the perception of each event's context in a straightforward manner. Each event is hosted inside a box initially displaying its title and time of occurrence. Upon selection, the event is scaled up and brought to the center of the display, while its box is transformed to host additional information at its sides [Fig 1, right]. The item displayed inside the box for detailed examination, according to its type (e.g. a video may be played and a 3D model may be rotated). Finally, when filters are applied through categories, the events are dimmed to indicate the fact that they are not available for interaction.

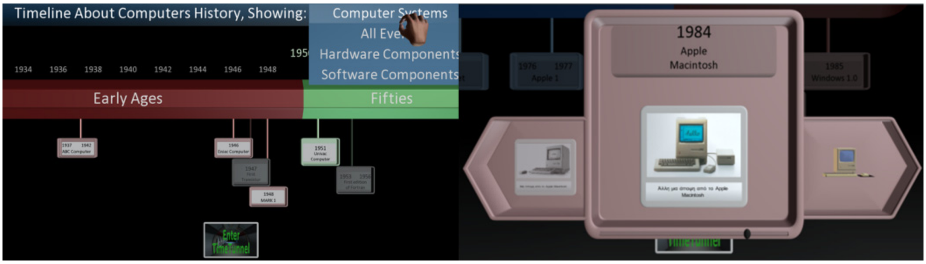

Fig. 1. The first view of the 3D timeline

The timeline's tunnel view (Fig 2) aims to provide an immersive view and focused on the sequential and exhaustive exploration of the presented information. The periods are hosted on the tunnel's roof sides while the events are placed on the floor. Aside each event the tunnel contains a "cavity" with a glass showcase, displaying a representative item that provides a visual clue. The showcase has a handle which the users may drag outwards to extend it, providing space to display additional multimedia content. The showcase may be again collapsed by dragging the hand inwards.

Fig. 2. The tunnel view

3.2 Interaction Methods

Touch Screen. Touch interaction involves the procedure of the user touching a point on the screen and the system projecting it to the virtual world. The available actions that the users may perform include clicking, dwelling upon an item and dragging.

Clicking or dwelling upon an item is translated by the system as a selection, whereas dragging serves a dual purpose: in the case where the user starts dragging an item, the action is interpreted as scrolling it in the corresponding direction, whereas if not, the system translates it as an intention to navigate in the virtual world.

Tangible Interaction. Another field of multimodal interaction is the application of tangible means using smart objects. In order to experiment with such items, a box was created, equipped with a 3-axes accelerometer which has the ability to transmit its orientation wirelessly. The smart box was employed as a joystick to navigate in the virtual world, using the rotation in two axes to navigate forward/backward and left/right. Additionally, the box was used for the rotation of the 3D models.

Kinesthetic Interaction. Kinesthetic interaction may be thought as “a unifying concept for describing the body in motion as a foundation for designing interactive systems” [3]. The types of kinesthetic interaction that this paper addresses include user’s position, controlling a virtual cursor, hand and leg gesturing. Leg gesturing and moving in space cover the need for spatial navigation in the virtual 3D world, whereas using hands may subsumed in the conceptual process of interacting with the elements displayed. The users may extend their hand towards the display in order to interact with the system. The movement of the hand is tracked and the user controls the virtual cursor while the hand is raised allowing the selection of visualized elements by placing it over an item for a short duration.

Hand gesturing was also applied as an additional technique that suits better tasks such as scrolling through successive elements (e.g. the elements comprising an event’s information). Furthermore, both hands are used in combination to simulate the process of pulling an item near or pushing it away in a natural and human-centric manner. The gestures may additionally involve continuous gestures that cause the system to respond while the users perform them. Such an example is the gesture for rotating the camera around the vertical axis of the virtual 3D world, which in the proposed system turns to the side while the users have their hand raised in the analogous direction (left/right). Leg gestures are applied by stepping towards any direction (right, left, up and down). Stepping is also continuous and causes the user (i.e. the virtual camera in the 3D world) to move towards the specified direction.

The interaction techniques used for the manipulation of the system should be robust and tolerant to possible user behavior that does not match the exact system specifications: the system should be able to prevent reacting in such a way that may be unexpected by the user, even at the cost of providing reduced yet sufficient functionality. Thus, the system adopts the concept of the user being able only to make a left swipe gesture with the right hand and a right swipe with the left hand.

4 Evaluation

The goal of the evaluation process was twofold. One the one hand, it aimed at assessing whether the users were more successful in using one interaction mode versus the other while trying to complete common tasks. On the other hand, it aimed at assessing the overall user experience of using each interaction mode. For the purpose of the first goal, Jakob Nielsen's User Success Rate method [10] was used. This method is good for comparison analysis and it is a simple yet effective way to estimate how successful the users were in using the system. For this method, the users were given a series of tasks to complete and each task was then marked as "Success" if the user was able to complete it in the first or second trial without asking for assistance, "Partial Success" if the user managed to complete the task after the third trial or after receiving minimum assistance by the facilitator, and as "Failure" if it took more than 3 trials to complete the task or if the user needed a lot of assistance in completing it. A simple formula was then used to calculate the Total Success Rate of the system. In order to assess the overall usability and user experience, the Think-Aloud process [9] was used during the evaluation, in which the participants were requested to express verbally their thoughts, comments, suggestions, and opinions throughout the completion of each task. In addition, at the end of the evaluation each user was asked to fill out a Likert scale based questionnaires for each interaction mode. The qualitative analysis has been presented in another paper [2]. This paper focuses more on the comparison of the user success rates of the two interaction modes.

Given that users were already familiar with touch screen technology and not so much with gestural interaction, it was expected that the touch screen interaction would show a higher user success rate than the gestures. Indeed, the quantitative results showed a slightly higher user success rate for the touch screen interaction than the gestural interaction, however, upon further examination of the qualitative data, the slight difference was found to have been caused by an interface design issue and not because of the interaction mode. These results are analyzed in more detail in the sections that follow.

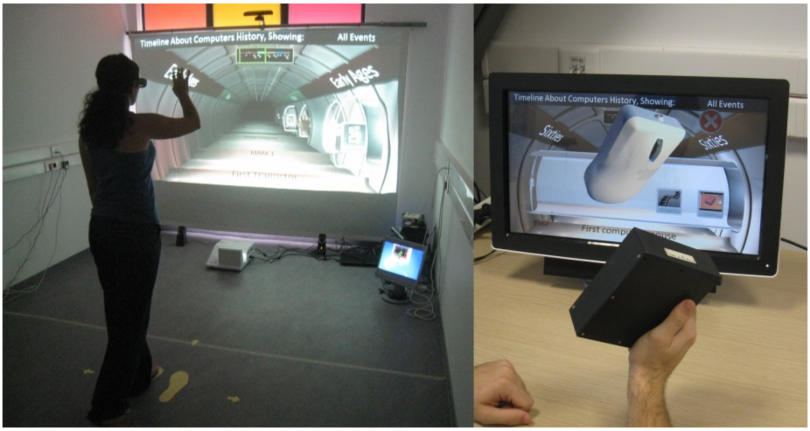

Fig. 3. The setups of the different evaluation segments

4.1 The Evaluation Process

A total of 16 volunteers participated in the evaluation, 7 females and 9 males from 20 to 40 years old. Twelve of the users (75%) had intermediate or high computer expertise whereas the other participants had limited expertise. Even though the majority of the users were familiar with computers and touch screen systems, they had very little to no experience in interacting with a system with gesturing. Two different evaluation set-ups were used (see Fig. 3 above) in the experiment. The first set-up was in a regular office room where the user would sit in front of a touch screen system and the second set-up was in a room where the application was projected on the wall and the user would stand in front of it and interact with hand and leg gestures (using Microsoft's Kinect). To eliminate bias towards either interaction method, the experiment was divided into two rounds and the users into two groups, where the first group started the interaction with the touch screen first in the office set-up and then with the kinesthetic interaction mode in the second set-up, while the second group started with the kinesthetic interaction first and then with the touch screen in the first set-up. The tasks of the second round were slightly different than the first round since the users already knew some of the answers from their first round of exposure to the application (see Tables 1 and 2).

Table 1. User Tasks performed either with the use of the touch screen or with hand and leg gestures

Task 1a	Navigate in the Timeline
Task 1b	Find how many decades are depicted in the Timeline
Task 1c	Find how many events are covered in the 1960s
Task 1d	Tell us which category of content you are currently viewing and how many other content categories exist
Task 2a	Zoom in the 1980s and find information relating to the Macintosh system
Task 2b	Find what kind of content is available for the above event
Task 3a	Find and select the photo "Another view of Apple Macintosh"
Task 3b	Open the photo and zoom in enough to read the name of the person depicted in the photo
Task 4	Find any video file, open it and play its content
Task 5	Find a file with a 3D model and explore it
Task 6	Use the SmartBox to interact with the same 3D model (the group that started with the gestures they were asked to interact with the model with hand gestures instead of the Smart Box in this particular task).

Table 2. User Tasks performed with the interaction mode that was not used in the previous round

Task 1	Enter in the Timeline and explore it
Task 2	Zoom in the 1960s and find information in relation to the first mouse device. What kind of content is available?
Task 3	Select one of the available photos and zoom in the face of the person depicted in it

Table 2. *(Continued)*

Task 4	Find a video file and play its content
Task 5	Zoom in the 1980s and find information relating to the Macintosh system
Task 6	Find what kind of content is available for the above event
Task 3a	Find and select the photo “Another view of Apple Macintosh”
Task 3b	Open the photo and zoom in enough to read the name of the person depicted in the photo
Task 4	Find any video file, open it and play its content
Task 5	Find a file with a 3D model and explore it
Task 6	Navigate to a different decade (either with the SmartBox or with leg gestures)

4.2 Result Analysis

User Success Rate. In the first round of the evaluation, the first group of users completed 11 tasks using the touch screen while the second group of users performed the same 11 tasks but with gestures. From this round of evaluation, the User Success Rate of the touch screen interaction was 89% and the User Success Rate of the gestural interaction was 82% (see fig. 4 and 5). In the second round the users from the first group performed similar tasks, but with gestures this time and the users from the second group performed these same tasks with the touch screen. This way both groups performed the same tasks using both interaction modes. The User Success Rate for the touch screen of the second round was 93,5% and the User Success Rate for the gestural interaction mode was 91% (see fig 6 and 7). As the results show both interaction modes produced higher User Success Rates in the second round of the experiment, but this was expected since the users were now more familiar with the interface and the system itself.

Combining the User Success Rates for each interaction mode from both rounds of the experiments produced a final **Total User Success Rate of 91%** for the touch screen interaction mode and a final **Total User Success Rate of 86%** for the gestural interaction mode.

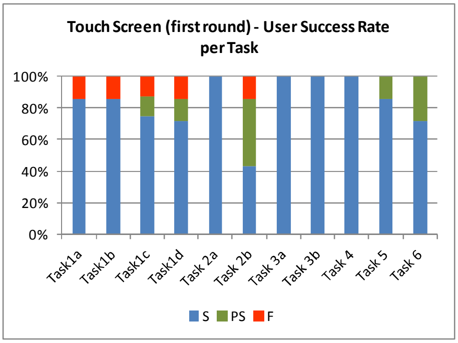

Fig. 4. Touch Screen User Success Rate per Task (first round)

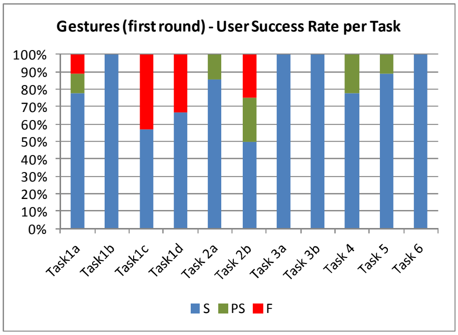

Fig. 5. Gestures User Success Rate per Task (first round)

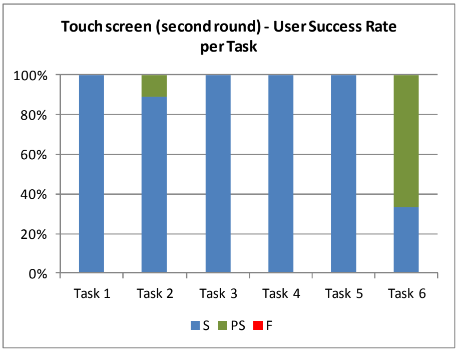

Fig. 6. Touch Screen User Success Rate per Task (second round)

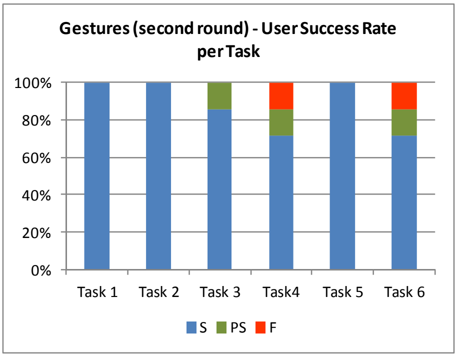

Fig. 7. Gestures User Success Rate per Task (second round)

Both interaction modes received a very high user success rate, which proves that they were effective and efficient methods for interacting with the system. The high user success rates were also supported by the qualitative analysis of the observations and the Think Aloud process. From the observations during the experiment, it was shown that the majority of the users understood quickly the Timeline concept and how the information was structured in it and thus the users didn't exhibit any instances of feeling lost in the application either in the 2D representation or the 3D Tunnel representation of the Timeline.

Touchscreen Interaction and Tangible Smart Box. In the touch screen interaction mode, the users instantly knew how to perform most of the functions such as zooming in and out, dragging, scrolling, and selecting an object without receiving any particular instructions from the facilitator. The touch screen gestures used in the application seemed to fit the mental model of the users that had extensive touch screen experience. One of the main comments that came up repeatedly from the users through the evaluations was the lack of multi-touch capabilities, which a lot of the users suggested adding it as a feature. Some of the users also expressed preference of controlling the zoom level by dragging the slider instead of just pressing the (+) and (-) buttons and iterating through multimedia content by dragging the slider instead of just the content elements. The SmartBox that was used as a complimentary method for the users to navigate in the Timeline and to manipulate the 3D model in Task 6 received mixed reactions from the users and produced a few Partial Success results in the evaluation. Overall, the use of the Smart Object as a complimentary way of the touch screen interaction was not as well accepted by the users, who found that having to switch from touch screen mode to the smart box was adding unnecessary burden to the user's interaction experience. Furthermore, users commented on the need for a direct one-to-one representation of the virtual model with the Smart Box, which is not currently supported.

Kinesthetic Interaction (hand/legs Gestures). The users also didn't exhibit any serious problems in interacting with the application with hand and leg gestures even though they were much less familiar with this interaction mode than with the touch screen. The supported gestures were easy for them to use and representative of the function they supported. Additionally, the segmentation of the interaction process into two distinct categories (leg gesturing for navigation in the 3D Tunnel representation, hand gestures and the virtual cursor for interaction with the visualized elements) provided a straightforward conceptual model of how to interact with the system for them. Tasks 1c and 1d were the tasks that caused the most failures in the kinesthetic interaction mode, but that was more due to a design and display issue, than a problem of the interaction mode itself. Some of the graphics of the design did not stand out as much in the second set-up with the projection on the wall and were thus missed by the users that started the evaluation in that set-up. A few of the users suggested that the movement required for some of the gestures such as pulling or pushing items using both hands, should be shorter in order to eliminate the fatigue factor setting in after prolonged interaction with the system. Leg gesturing was almost unanimously accepted as a complimentary way of navigating through the 3D Tunnel model of the application and only one user could not navigate using his legs because Kinect failed to successfully recognize the exact placement of his legs due to the trousers the user was wearing. The combination of stepping in any direction in order to travel in space and interacting with the system elements using the hands proved a powerful method which well received by the users. This observation is more evident in the non-expert users, who supported the leg gestures even more than the expert users, as they felt more comfortable with handling the system naturally, but in a strictly defined manner. Overall, the participants found the conceptual model of moving in the space to be efficient, tireless and fascinating.

5 Conclusions

The selection of the appropriate interaction technique relies on the purpose of each system and the context in which it is designed to be used. Touch interaction proved to provide a complete and thorough set of instructions to manipulate 3D environments, suitable mainly for everyday use. Tangible artifacts received mixed reactions and their usefulness was questioned mainly due to their inability to fully handle a complex environment in terms of both navigation and object selection. Remote system handling through non-instrumented user tracking provided rich interaction vocabulary, which the users are able to successfully memorize and use in combination. The fun factor of expressing themselves by making human-like movements overruled the tiredness that appeared after extensive use, making kinematic interaction suitable for setups which involve entertainment rather than serious work.

Acknowledgements. The work reported in this paper has been conducted in the context of the AmI Programme of the Institute of Computer Science of the Foundation for Research and Technology-Hellas (FORTH).

References

1. Alexander, J., Han, T., Judd, W., Irani, P., Subramanian, S.: Putting your best foot forward: investigating real-world mappings for foot-based gestures. In: Proceedings of the 2012 ACM Annual Conference on Human Factors in Computing Systems, pp. 1229–1238. ACM (May 2012)
2. Drossis, G., Grammenos, D., Adami, I., Stephanidis, C.: To appear in INTERACT 2013. Springer (2013)
3. Fogtmann, M.H., Fritsch, J., Kortbek, K.J.: Kinesthetic interaction: revealing the bodily potential in interaction design. In: Proceedings of the 20th Australasian Conference on Computer-Human Interaction: Designing for Habitus and Habitat, pp. 89–96. ACM (December 2008)
4. Grammenos, D., Zabulis, X., Michel, D., Sarmis, T., Tzavanidis, K., Argyros, A., Stephanidis, C.: Design and Development of Four Prototype Interactive Edutainment Exhibits for Museums (2011)
5. Hilliges, O., Izadi, S., Wilson, A., Hodges, S., Mendozam, A.G., Butz, A.: Interactions in the air: adding further depth to interactive tabletop. In: Proc. UIST 2009, pp. 139–148. ACM (2009)
6. Jaimes, A., Sebe, N.: Multimodal human–computer interaction: A survey. Journal Computer Vision and Image Understanding Archive 108(1-2), 116–134 (2007)
7. Molyneaux, D., Gellersen, H.: Projected interfaces: enabling serendipitous interaction with smart tangible objects. In: Proc. TEI 2009, pp. 385–392. ACM (2009)
8. Nickel, K., Stiefelhagen, R.: Pointing Gesture Recognition based on 3D-Tracking of Face, Hands and Head Orientation. In: Proc. ICMI 2003, pp. 140–146. ACM (2003)
9. Nielsen, J.: Evaluating the thinking-aloud technique for use by computer scientists (1992)
10. Nielsen, J.: Success Rate: The Simplest Usability Metric (2001), `http://www.nngroup.com/articles/success-rate-the-simplest-usability-metric/`
11. Oviatt, S.L.: Advances in Robust Multimodal Interface Design (2003)
12. Pakkanen, T., Raisamo, R.: Appropriateness of foot interaction for non-accurate spatial tasks. In: CHI 2004 Extended Abstracts on Human Factors in Computing Systems, pp. 1123–1126. ACM (April 2004)
13. Papadopoulos, C., Sugarman, D., Kaufmant, A.: NuNav3D: A Touch-less, Body-driven Interface for 3D Navigation (39) (2012)
14. Sangsuriyachot, N., Mi, H., Sugimoto, M.: Novel Interaction Techniques by Combining Hand and Foot Gestures on Tabletop Environments. In: Proc. ITS 2011, pp. 268–269. ACM (2011)
15. Terrenghi, L., Kranz, M., Holleis, P., Schmidt, A.: A cube to learn: a tangible user interface for the design of a learning appliance. Journal Personal and Ubiquitous Computing Archive 10(2-3), 153–158 (2005)
16. Valkov, D., Steinicke, F., Bruder, B., Hinrichs, K.: Traveling in 3D Virtual Environments with foot gestures and a multitouch enabled World In Miniature (2010)
17. Yoo, B., Han, J.J., Choi, C., Ryu, H.S., Park, D.S., Kim, C.Y.: 3D remote interface for smart displays. In: Proceedings of the 2011 Annual Conference Extended Abstracts on Human Factors in Computing Systems, pp. 551–560. ACM (May 2011)