Article

On the Use of Mobile Devices as Controllers for First-Person Navigation in Public Installations †

Spyros Vosinakis * and **Anna Gardeli**

Department of Product & Systems Design Engineering, University of the Aegean, 84100 Syros, Greece
* Correspondence: spyrosv@aegean.gr; Tel.: +30-22-8109-7125
† This paper is an extended version of our presentation in the 5th International Conference AVR 2018, Otranto, Italy, 24–27 June 2018.

check for updates

Abstract: User navigation in public installations displaying 3D content is mostly supported by mid-air interactions using motion sensors, such as Microsoft Kinect. On the other hand, smartphones have been used as external controllers of large-screen installations or game environments, and they may also be effective in supporting 3D navigations. This paper aims to examine whether a smartphone-based control is a reliable alternative to mid-air interaction for four degrees of freedom (4-DOF) fist-person navigation, and to discover suitable interaction techniques for a smartphone controller. For this purpose, we setup two studies: A comparative study between smartphone-based and Kinect-based navigation, and a gesture elicitation study to collect user preferences and intentions regarding 3D navigation methods using a smartphone. The results of the first study were encouraging, as users with smartphone input performed at least as good as with Kinect and most of them preferred it as a means of control, whilst the second study produced a number of noteworthy results regarding proposed user gestures and their stance towards using a mobile phone for 3D navigation.

Keywords: 3D navigation; large screens; smartphone; Kinect; mid-air interactions; user study

1. Introduction

Advances in imaging technology and natural user interfaces in the last decade have allowed the enhancement of public spaces with interactive installations that inform, educate, or entertain visitors using rich media content [1]. These digital applications are usually displayed on large surfaces placed indoors or outdoors, e.g., on walls, screens, or tables using projection or touch screen displays, and users can explore and interact with them in a natural, intuitive, and playful way. In many cases, these installations include 3D content to be explored, such as complicated geometric models or even large virtual environments. Most installations displaying 3D content are currently found in museums and cultural institutions [2], where the use of detailed geometry is necessary for the presentation of cultural heritage, e.g., by exhibiting virtual restorations of ancient settlements, buildings, or monuments. Other application areas of installations based on 3D content include interacting with works of art [3], exploring geographical [4] or historical [5] information, performing presentations [6], interacting with scientific visualizations [7], navigating in virtual city models for urban planning [8], etc.

The design of 3D interaction techniques for installations requires special attention in order to be intuitive and easy to perform by the general public. 3D environments introduce a different metaphor and extra degrees of freedom, and new users can easily get frustrated through repeated ineffective interactions [9]. The most fundamental, yet complicated interaction technique for any type of environment is user navigation. 3D spaces naturally require frequent movements and viewpoint changes in order to be able to browse the content from different angles, to uncover occluded parts of the scene, to travel to distant parts, and to be able to interact with objects from a certain proximity.

Navigation is a mentally demanding process for inexperienced users, because it involves continuous steering of the virtual body, as well as wayfinding abilities. As far as steering is concerned, a difficult challenge for designers is the meaningful translation of the input device into respective movements in the 3D world [10]. In most public installations with 3D content, user navigation takes place from a first-person point of view and involves the virtual walkthrough of interior and exterior spaces.

Typical desktop or multitouch approaches are not the most appropriate means of navigating and interacting with 3D content in public installations. A common setup in such systems is to present the content on a usually large vertical surface and to let visitors interact in a standing position at some distance, being able to look at the whole screen. The use of a keyboard or mouse is not very helpful for a standing user, whilst touch or multitouch gestures cannot be performed if visitors are interacting from a distance. To overcome these issues, public installations have been using solutions 'beyond the desktop,' usually based on natural user interfaces. Initially, the interaction techniques involved handheld controllers, such as WiiMote, or other custom devices, e.g., [11]. However, the use of handheld devices in public settings raises concerns about security and maintenance. More recently, users have been able to interact with 3D content in public settings using body gestures, without the need of any additional handheld or wearable device. Developers have taken advantage of low-cost vision and depth sensing technology and have created interactive applications, in which users can navigate or manipulate objects of a 3D scene using body movements and arm gestures in mid-air.

A variety of sensors have been used for mid-air interactions in public installations, the most popular one being Microsoft Kinect. Kinect can detect the body motion of up to four users in real time and translate them into respective actions. As such, it is appropriate for standing users navigating and interacting with a 3D scene from a distance, and has been already deployed in public museum environments, e.g., [12,13]. A secondary, less common option is the Leap Motion controller, which is considered faster and more accurate but is limited to hands-only interaction. The controller has to be at a near distance from the users' hands and it is therefore more appropriate for seated users, which is somehow limiting for public installations. Numerous techniques for first-person navigation have been implemented using the Kinect sensor, such as leaning the body forwards or backwards to move [14,15], rotating the shoulders to change direction [14–16], walking in place [15,16], using hands to indicate navigation speed and direction [17,18], using both hands to steer an invisible bike [18], etc. Two comparative studies have also been setup to assess the effectiveness and usability of Kinect navigation techniques in field or laboratory settings [15,16], identifying preferences and drawbacks of the aforementioned techniques.

Most evaluations of interactive 3D installations using Kinect have concluded that it is a motivating and playful approach but not without problems. Some people feel embarrassed to make awkward body postures or gestures in public [19]. Also, there are users who find the interactions tiring after a while because of the fatigue caused by some gestures, e.g., having to hold arms up for a long time. Finally, the presence of other people near the installation may cause interference to the sensor and therefore most of these installations require that an area near the user is clear from visitors. An alternative to mid-air interactions for navigating in public installations that has been recently proposed is the use of a mobile device as a controller [20]. Most people carry a modern mobile device (smartphone or tablet) with them with satisfactory processing and graphics capabilities and equipped with various sensors. Following the recent trend of "bring your own device" (BYOD) in museums and public institutions [21], where visitors use their own devices to access public services offered by the place, one could easily use her device for interacting with a public installation. For example, using the public WiFi, one could download and run a dedicated app or visit a page that turns her device into a navigation controller. The use of mobile devices as controllers has already been tested in other settings, e.g., games and virtual environments, with quite promising results [22–24]. This alternative may have some possible advantages compared to mid-air interactions. It can be more customizable, it can lead to more personalized experiences by tracking and remembering individual users, and it could also deliver custom content on their devices, e.g., a kind of 'reward' for completing a challenge.

The aims of this work are to examine whether a mobile device used as a controller can be a reliable solution for first-person 3D navigation in public installations, and to determine the main design features of such a controller. We carried out two successive studies for this purpose.

In the first study, we sought to explore whether a smartphone controller can perform at least as good as Kinect-based navigation, which is the most common approach today. We setup a comparative study between mid-air bodily interactions using Kinect and tilt-based interactions using a smartphone in two environments and respective scenarios: A small museum interior, in which the user has to closely observe the exhibits, and a large scene with buildings, rooms, and corridors, in which the user has to effectively navigate to selected targets. The interaction techniques used in this study have been selected and adapted based on the results of previous research, i.e., we used a mid-air interaction involving the leaning and rotation of the upper body, which generated the highest outcomes in [15] and was also one of the prevalent methods in [16], and a technique based on the tilting and rotation of the handheld device, which was also discovered as usable in [7,25,26]. A testbed environment developed for the study automatically measured the time spent to complete each scenario, the path travelled, the number of collisions, and the total collision time. Furthermore, subjective ratings and comments for each interaction technique were collected by the users through questionnaires and follow-up discussions. The results of the first study indicated that the smartphone performed at least as good as Kinect in terms of usability and performance, and it was the preferred interaction method for most of the participants.

Following the encouraging results of the first study, we aimed to look in more depth at the interaction techniques to be used for the design of a mobile controller. For this purpose, we setup a gesture elicitation study to collect preferred gestures from users and improve the guessability of the designed interactions [27]. We had our participants propose their own gestures for a series of navigation actions: Walking forward and backwards, rotating to the left or right, looking up or down, and walking sideways. They were free to select between (multi-)touch actions, rotating or moving the whole device, or a combination of them, and they could propose any visual interface on the device. Whenever they proposed a gesture, we tested it in the museum environment of the first study using a Wizzard of Oz technique and had our users reflect about it. The results of the study led to interesting observations regarding the preferred gestures of users and the different ways in which users mapped mobile actions to 3D movements in the projected environment.

We present the results of our studies and a discussion about their implication for the design of novel interaction techniques for virtual reality applications presented on public displays.

2. Materials and Methods

Initially, we setup a comparative study between Kinect-based and smartphone-based interaction techniques for first-person navigation in 3D environments. Although the aim was to compare these two modalities, we decided to include keyboard input as a third modality in the study as well. The reason for this decision is first because the keyboard is a common input method for users, so it can be used as basis for comparison, and second because this input helped users to familiarize with the scenes in the scenarios, before trying the other two modalities. Furthermore, we decided to use four degrees of freedom (4DOF) for first person navigation instead of the two or three used in most of the other studies, because the extra degrees (looking up or down and walking sideways) are useful for architectural walkthroughs and virtual museums. The study focused on the perceived usability and performance of the input modalities in two different scenarios.

After the results of the first study that were generally in favor of the mobile device, we conducted a follow-up gesture elicitation study focusing solely on the smartphone control. Our aim was to address the following question: How do users contemplate interacting with public displays, using smartphone-based gestures for navigating in 3D environments?

2.1. Interaction Techniques

The navigation techniques we designed for the input modalities of the comparative study follow the same concept. The user can use a special action to switch between three navigation modes, each of which has two degrees of freedom. The available navigation modes are the following:

- **Walk/turn**: Move forwards or backwards and turn to the left or right;
- **Walk sideways (strafe):** Move sideways; and
- **Look around**: Look up or down and turn to the left or right.

The interaction techniques for the three modalities (keyboard, Kinect, and smartphone) in each navigation mode (shown in Figure 1) are the following:
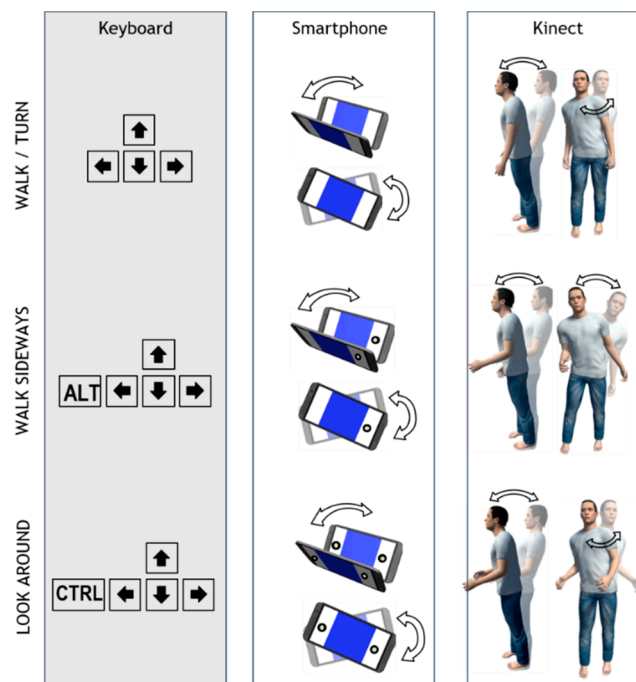

Figure 1. The interaction techniques for the three input modalities (keyboard, Kinect, smartphone) regarding the respective navigation modes (walk/turn, walk sideways, look around).

For the 'walk/turn' navigation mode:

- The cursor keys of the keyboard are used;
- The user's body is leaning forwards and backwards, while her shoulders have to rotate left and right in order to turn, for Kinect input;
- The smartphone device must be tilted forwards or backwards to move to that direction and rotated like a steering wheel to turn left or right, while held by both hands in a horizontal direction (landscape).

For the 'walk sideways' (strafe):

- The ALT key of the keyboard combined with the cursor keys is used;
- The user's one arm (either left or right) is raised slightly by bending her elbow and the user's body is leaning left and right to move to the respective side, for Kinect input;
- On the smartphone device, one button should be pressed (either left or right, as both edges of the screen work as buttons) and the device must be rotated like a steering wheel to walk sideways left or right.

For the 'look around' mode:

- The CTRL key of the keyboard combined with the cursor keys is used: Up and down is used to move the viewpoint upwards or downwards, respectively, and left and right to rotate it;
- Both of the user's arms are raised slightly, while the user's body leans forwards or backwards to look down or up, respectively, and rotates her shoulders to turn to the left or right;
- The user presses both buttons of the smartphone device and tilts or rotates the device to turn the view to that direction.

A small pilot study was set up to calibrate the testbed environment prior to the main study. Four users participated in order to adjust the first-person controller movement and rotation velocity based on their feedback. They were asked to suggest any changes to the navigation and rotation velocity after some familiarization with the navigation techniques. The process was repeated until the users felt more at ease with the interface, while the speeds were recalibrated on the fly. All users achieved comparable required values, which were considerably slower from the testbed's original values. Based on these outcomes, the testbed environment's movement and rotation speeds were adapted to nearly 60% of their initial value.

Concerning the follow-up elicitation study, we intended to elicit alternative gestures and interaction techniques for each navigation mode regarding the smartphone-based modality using the same 4DOF. Therefore, we defined the following distinct tasks, one for each DOF:

- **Walk fwd-back:** Walk forwards or backwards;
- **Rotate left-right**: Rotate the viewpoint to the left or right;
- **Look up-down:** Rotate the viewpoint upwards or downwards; and
- **Walk sideways**: Walk to the left or right (strafe) without turning the viewing direction.

2.2. Equipment and Setting

A testbed environment was set up in Unity game engine to support navigation in 3D scenes and to record the user's behavior, such as the travel path, duration, and collision. The environment directly supported keyboard and Kinect input. Additionally, a first-person controller component was implemented as a smartphone app in order to translate the user's input into respective actions of the virtual body. It transmitted the rotation values and button presses to the testbed application through the WiFi network, and the virtual body moved accordingly.

Both studies took place in the laboratory and shared the same settings and equipment. For the testbed environment, a PC with Intel Xeon CPU 3.70 GHz, 16 Gb Ram (Intel, Santa Clara, CA, USA), and an NVIDIA Quadro K4200 graphics card (NVIDIA, Santa Clara, CA, USA) was utilized, and a projection screen through an Epson EB-X24 Projector (Epson, Nagano, Japan) was used to display the scene. For all three modalities in the comparative study and the one used for elicitation, each user was standing at a distance of about 2.5 meters from the projection screen. In the keyboard-based input, users were seated using a wireless keyboard; in the Kinect-based input, users were standing in front of a Microsoft Kinect 2 (Xbox One) sensor (Microsoft, Redmond, WA, USA); and in the smartphone-based input, they were standing at the same spot holding a Xiaomi mi4i phone that was running the controller app. Figure 2 shows the setup of the user study in terms of screen dimensions, sensor placement, and user distance from screen.
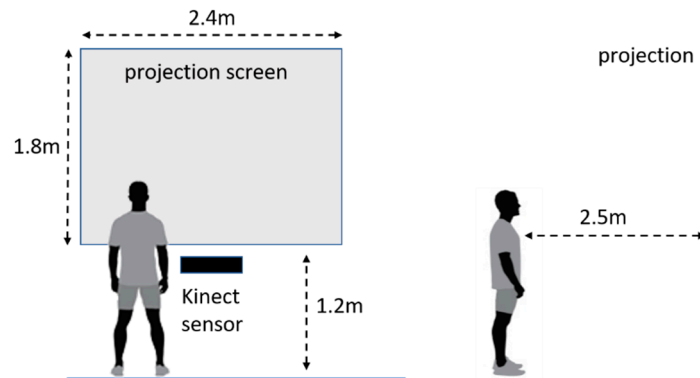
Figure 2. Setup of the experiment. Left: projection screen dimensions and placement and sensor placement. Right: distance between the user and projection screen.

In the testbed setting, three distinct scenes were prepared and used. These were:

- **Familiarization**: A simple scene that allows users to familiarize with each interaction technique. It displays a digitized version of the Stonehenge site.
- **Buildings**: An indoor and outdoor scene displaying abandoned buildings with rooms and corridors. Users' task was to walk around a building, and to carefully maneuver their virtual body through narrow doors.
- **Museum**: A small interior scene featuring a digitized version of the Hallwyl Museum Picture Gallery. Users' task was to walk through the hallways slowly and focus on particular displays.

Figure 3 presents a user interacting with the 3D environment using Kinect and a smartphone as input, and the buildings and museum scenes used in the study.
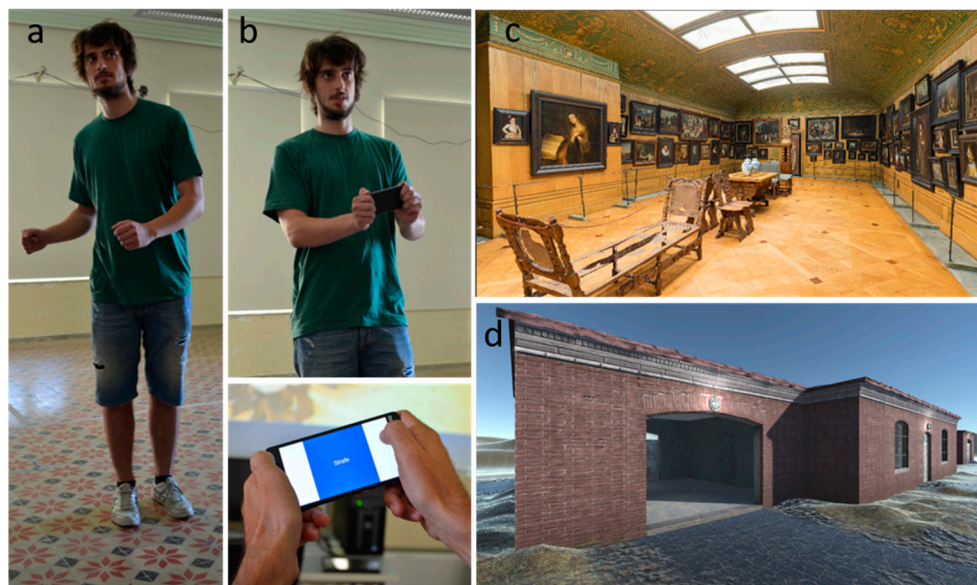

Figure 3. A user interacting with the 3D environment using Kinect (**a**) and a smartphone (**b**) and the two scenes used in the study: museum (**c**) and buildings (**d**).

The museum scene was also utilized in the elicitation study. However, in this case, they could not actively interact with it. They navigated through a Wizard of Oz technique. Wizard of Oz is a common technique for evaluating early prototypes in human–computer interaction studies, where users believe that they interact with a system, but in reality, a human is partially controlling the systems' response through observing user input. In our case, users performed the desired gestures and the evaluator

controlled the viewpoint using keyboard input, to produce the effect that their actions had actual impact on the movement.

2.3. Participants

In the comparative study, 22 users participated, 11 males and 11 females. The age span was wide (between 20 and 50), and most users were under 30 (M: 25.4, SD: 7.8). The experience of the respondents was quite balanced with computer games and 3D environments. In total, 11 users reported that they had large or very large experience, 6 had none or little experience, and 5 medium.

In the elicitation study, 28 users participated, 15 males and 13 females. There was also a broad age range (between 19 and 47), but most participants were under 25 (M: 24.6, SD: 6.27). Concerning the participants' experience with 3D computer games, the result was mixed, as 9 users reported that they had large or very large experience, 7 had none or little experience, and 12 medium.

The majority of the participants were students and faculty from our department in both studies. However, none of them took part in the both studies. Therefore, in the elicitation study, none of the users had prior experience with navigation in a 3D environment using a smartphone.

2.4. Procedure

2.4.1. Comparative Study

The study adopted a within-subjects design, so each subject used all three interaction modalities in both scenes, i.e., 6 trials per user. The procedure was as follows.

First, users were introduced to the research purpose and procedure and were requested to fill out their gender, age, and experience with games and 3D environments in an initial questionnaire.

Next, they had to use the three interaction modalities and complete the scenarios. All users began with keyboard input, followed by the other two modalities, but with an alternating order, to counterbalance potential order effects.

The users had to navigate through all three scenes for each interaction modality. They were initially put in the scene of familiarization and allowed to navigate around using the interface until they felt comfortable with it. Then, they went on to the buildings scene, where they had to carry out a particular task: They had to enter a building and move in three particular rooms to assigned positions.

Third, they had to move around the museum scene and focus on four particular exhibits by attempting to bring them to the middle of the screen. In both scenarios, the target positions were shown to the users by the evaluators during their navigation, so the actual challenge was to steer their virtual body to the designated targets. At the end of each modality, users filled out a questionnaire for the interface and interaction technique with their subjective ratings and remarks.

Finally, users were asked to choose between navigation based on Kinect and a smartphone as their preferred technique and make any conclusive remarks. Each user session took about 35 min to complete.

2.4.2. Gesture Elicitation Study

In the follow-up gesture elicitation study, we used the following procedure.

First, users were introduced to the purpose and procedure of this study and filled in their gender, age, and experience with 3D computer games.

Then, they were given a switched off mobile device to hold and were asked to use it to propose gestures for the navigation tasks, in a way that seems more suitable/appropriate to them. We explained to them that they were free to use any possible action on the mobile phone, whether on the surface (touch/multi-touch actions) or based on motion (move or tilt actions), and they could propose any possible visual or interactive element to appear on the device's screen. Also, they were free to hold the phone and place their fingers in any desired way.

Participants had to perform their proposed gestures for each of the navigation tasks in this sequence: Walk fwd-back, rotate left-right, look up-down, and walk sideways. For each task, we used the following protocol: First, we performed the action on the environment (using the keyboard input) and we asked the user to think of a suitable phone gesture to cause that effect. When the user was ready, she performed and explained the gesture (and the desired effects), and we tested it together on the environment using the Wizard of Oz technique. This helped users reflect on their choice and possibly even propose minor alterations. After that, the user was asked to provide her subjective ratings regarding the suggested gesture based on the questions, 'how easy was it for you to produce this gesture' and 'how appropriate is your gesture for the task.' These questions were adapted from the study presented in [28].

During the study, participants were also asked to combine some of their proposed gestures for reflection and reconsideration. Again, we used the Wizard of Oz technique to test the combinations. The combined actions were the following: (a) Steering the virtual body by combining walk fwd-back with rotate left-right and (b) looking around by combining look up-down with rotate left-right. We tested the first combination after the rotate left-right gesture was proposed, and the second after the look up-down gesture. As expected, sometimes there were conflicts between proposed gestures or unusable combinations, and we discussed with users about their possible resolution.

Finally, users were asked to give any conclusive comments. Each session lasted approximately 20 min.

2.5. Collected Data

In both studies, we collected qualitative and quantitative data, both regarding each interaction technique (comparative study) and users' intuitive suggestions concerning smartphone-based gestures for each of the navigation tasks (elicitation study).

In the first study, users entered their subjective scores in a 5-point Likert scale regarding: (a) Ease of use, (b) learnability, (c) satisfaction, (d) comfort, and (e) accuracy (see Appendix A). The questions were chosen and adjusted from popular usability questionnaires. In addition, through the testbed environment, the following data were gathered for each task: (a) Task's completion time, (b) number of collisions, (c) total collision time, and (d) path. The user's collision with a wall or obstacle during navigation was deemed an error, so we decided to include both the amount of collisions and the total time the user was colliding (e.g., while sliding on a wall) as an indication of ineffective steering. The recorded route can provide a qualitative summary of each user's navigation quality and allow the complete distance traveled to be calculated as well. Eventually, users were asked to pick the input modality they preferred, between Kinect and a smartphone.

In the second study, users gave their subjective scores regarding easiness and appropriateness of each proposed gesture using a 5-point Likert scale (see Appendix A). Moreover, the responses of users were recorded both in video and in writing. The recorded videos provided us with more insights concerning the way users preferred to hold and interact with the device, their reflection and reconsideration of suggested gestures, and more qualitative information of each user's thoughts and comments regarding the process.

3. Results

3.1. Comparative Study between Smartphone-Based and Kinect-Based 3D Navigation

The gathered data were evaluated at the end of the experiment to provide insight into the usability and performance of each modality of interaction. For each of the gathered measures, we calculated mean values, standard deviations, and confidence intervals of 9%. Using one-way variance analysis (ANOVA) to check for statistically significant variations between the three modalities, the time, collisions, and distance information were further evaluated and post-hoc Tuckey HSD experiments for pairwise comparisons were conducted where necessary. The post-hoc experiments were used

whenever statistical significance was detected; in that case, all possible pairs of interaction techniques were tested to detect which of the techniques had a significant difference. Because of the use of the Likert-scale, the subjective ratings were non-parametric, so they were further analyzed using pairwise Wilcoxon signed-rank tests for statistically significant variations.

3.1.1. Time, Collisions, and Distance Travelled

Figure 4 shows the mean values and 95% confidence intervals for task time (seconds), amount of collisions, collision time (seconds), and traveled distance (meters) for the tree input modalities.
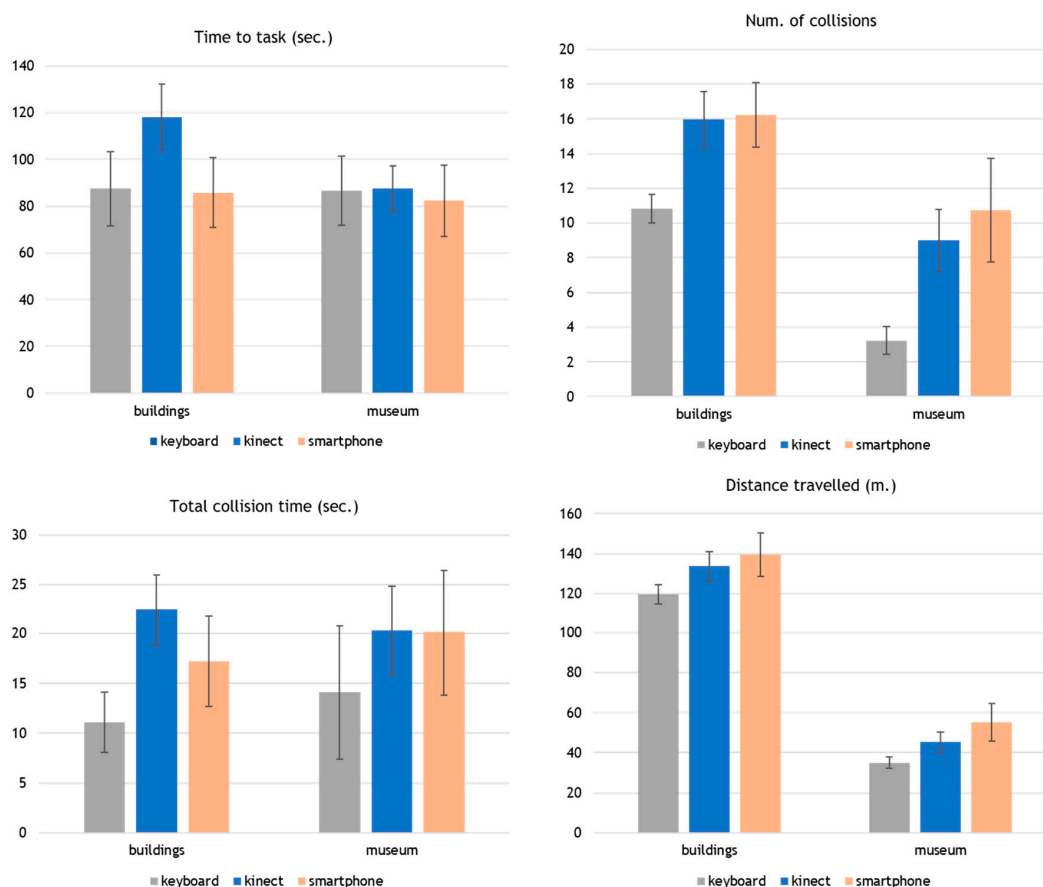

Figure 4. Mean values of the total time to finish the task in seconds (upper left), number of collisions (upper right), total collision time in seconds (lower left), and total distance travelled in meters (lower right) in the buildings and museum scenes. Error bars represent 95% confidence intervals.

There was a statistically significant difference between the three modalities of input in the building scene (F(2.63) = 5.44, p = 0.007) and no statistically significant difference in the museum scene with respect to the total time to task. Post-hoc Tuckey HSD tests disclosed that the total keyboard time was considerably smaller in the building scene compared to Kinect (p = 0.02) and that the total smartphone time was also considerably smaller compared to Kinect (p = 0.013).

The keyboard exceeded the other two modalities in the total amount of collisions. In particular, for the buildings (F(2.63) = 15.03, p < 0.001) and the museum (F(2.63) = 13.33, p < 0.001), the variance analysis indicated a statistically significant difference. Post-hoc Tuckey HSD tests showed that in both scenes, the number of collisions using the keyboard were significantly lower compared to Kinect (p = 0.001 in both scenes) and the smartphone (p = 0.001 in both scenes).

Regarding the total collision time, there was a significant difference between the three modalities in the buildings scene (F(2.63) = 8.34, p < 0.001) and no statistically significant difference in the museum scene. The post-hoc Tuckey HSD tests for the collision time in the buildings scene showed that when

using the keyboard the total collision time was significantly smaller compared to Kinect ($p = 0.001$). No statistically significant difference was found between the keyboard and smartphone.

Finally, for the total distance travelled, the ANOVA tests revealed statistically significant differences in both scenes (buildings scene: $F_{(2.63)} = 5.86$, $p = 0.004$, museum scene: $F_{(2.63)} = 9.50$, $p < 0.001$). The post-hoc analysis in both scenes revealed that when using the keyboard, users travelled significantly less distance compared to using the smartphone (buildings scene: $p = 0.004$, museum scene: $p = 0.001$). No statistically significant difference was found between the keyboard and Kinect.

3.1.2. Subjective Ratings

Figure 5 summarizes the outcomes of the 5-point Likert questions for the subjective measures in each interaction modality with averages and confidence intervals of 95%.
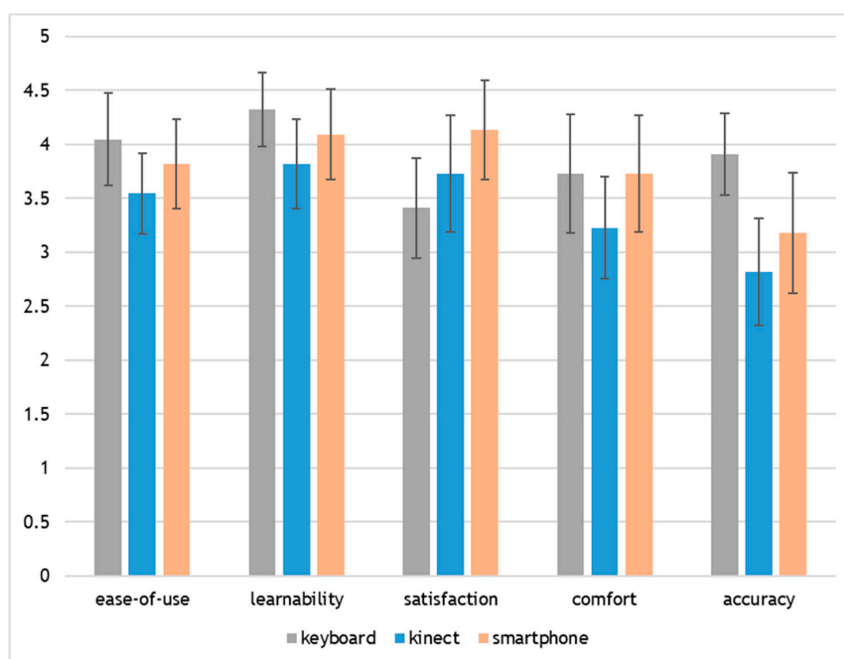

Figure 5. Mean values of subjective user ratings regarding ease-of-use, learnability, satisfaction, comfort, and accuracy. Error bars represent 95% confidence intervals.

The analysis using Wilcoxon signed-rank tests identified some significant differences among the modalities in terms of satisfaction, comfort, and accuracy. Specifically, in satisfaction, the smartphone scored significantly better than the keyboard ($p = 0.016$), and in accuracy, keyboard control was rated as significantly better than both other modalities ($p = 0.005$ for Kinect, $p = 0.041$ for smartphone). Furthermore, regarding comfort, users considered the smartphone significantly better than Kinect ($p = 0.003$).

Looking at the mean scores of the subjective ratings, it is obvious that the keyboard scored at least as good as the two other modalities, although it scored lower in satisfaction. More specifically, the mean scores regarding satisfaction were 3.40 (SD: 1.11) for the keyboard, 3.73 (SD: 1.29) for Kinect, and 4.14 (SD: 1.10) for the smartphone. The mean scores for the smartphone were slightly higher compared to those of Kinect in all ratings.

Finally, regarding the user's preference of input modality, the smartphone significantly exceeded Kinect. Seven users (31.8%) preferred Kinect, whilst 15 users (68.2%) preferred the smartphone-based input.

3.1.3. User Comments and Observations

On the basis of our observations and user comments during the study, it is apparent that users navigating with Kinect or the smartphone could not keep their virtual body completely inanimate,

unlike the keyboard input. We realized that some skilled users switched to 'look around' mode deliberately when they wanted to prevent the viewpoint motion. Also, some inexperienced users found the smartphone-based and Kinect-based modalities too sensitive. They were unfamiliar with the analog velocity control (i.e., changing proportionally to the amount of leaning or rotating), and they tended to make large movements, which in turn resulted in higher velocities and less sense of control. On the other side, it was much simpler for more skilled users to gain control of the techniques.

Some users reported that they had forgotten about the moving backwards option in Kinect-based or smartphone-based interaction, which explains why they did not use it. Also, the strafe movement was found to be not very helpful by a few users.

Regarding Kinect-based input, some users confused the shoulder rotation with side-ways leaning, leading to problems with virtual body steering.

As far as smartphone-based input is concerned, there were users who forgot to rotate it like a steering wheel, turning the device left or right. Also, a lot of users preferred diagonally holding the device (tilted to the front) as a 'resting' position, instead of vertically, as they claimed that this is the usual way they hold a smartphone. One user's comment was that he would prefer an interaction technique for one hand instead of two.

Finally, among the user groups, we noticed some patterns. Users who reported no game and 3D environment experience appeared to prefer the Kinect input, while experienced gamers were much more passionate about the smartphone. Furthermore, the preference for input modality among female users was almost equally split (6 for smartphone and 5 for Kinect), while the preference for smartphone-based input (9 for smartphone and 2 for Kinect) was much greater among male users.

3.2. Gesture Elicitation Study for Smartphone Control

We collected 112 gestures in total from the 28 participants. The gestures were studied and analyzed based on the recorded video footage, which lasted around 4 h and 30 min.

3.2.1. Taxonomy

To gain better insight into users' suggestions and preferences, we created a taxonomy to categorize the gestures based on various dimensions. These dimensions were based on previous taxonomies for mobile gestures and 3D interactions [29,30] and adapted to the needs of this study.

In total, our taxonomy uses six dimensions: Phone orientation, fingers used, preferred angle, manipulation, mapping, and gesture type. Table 1 summarizes the dimensions and values of the proposed taxonomy.

Table 1. The taxonomy used in the study.

Dimension	Values	Description
phone orientation	portrait	User is holding the phone in vertical (portrait) orientation
	landscape	User is holding the phone in horizontal (landscape) orientation
fingers used	zero fingers	No touch actions are performed
	one finger	A single touch or drag action is performed
	two fingers	An action that uses two fingers
preferred angle	horizontal	User is holding the phone almost horizontally—parallel to the ground
	diagonal	Holding the phone with an angle around 45° from the ground
	vertical	User is holding the phone almost vertically to the ground, facing the screen
manipulation	viewpoint	User actions manipulate the viewpoint (camera)
	world	User actions manipulate the whole scene
mapping	position	The gesture directly controls the position of the viewpoint
	velocity	The gesture controls the moving velocity of the viewpoint

Phone orientation describes the way in which users held the mobile phone. Some people held the phone with one hand in portrait mode, whilst others preferred to hold it with two hands in landscape mode.

The fingers used dimension indicates how many fingers were used to actively touch the phone surface to perform an action. Users performing motion gestures usually did not touch the surface (zero fingers), although there were a few exceptions where phone movement was combined with touch. In most touch gestures, users used a single finger (one finger), and in some cases, they used a second finger as well of the same or the other hand (two fingers).

The preferred angle is the way in which users tended to hold their phone with respect to the surface's angle from the ground. The possible values are horizontal, if the phone's surface is almost parallel to the ground, diagonal if it has an angle near 45° from the ground, and vertical if it is almost vertical to the ground. Most users kept the same angle for all gestures. In the case of motion gestures, we considered the preferred angle as the one that the users named as the 'resting pose'. We explicitly asked them for that.

The manipulation dimension has to do with the way in which their gestures were interpreted. Most users, especially experienced ones, proposed a direct mapping between their motion and the camera's motion, e.g., when they moved their finger to the left they expected the camera to rotate to the left as well. Following the taxonomy of [31], we called this metaphor viewpoint manipulation. On the other hand, there were some users who were more familiar with 2D interactions, such as flicking images when using their mobile phone, and they proposed gestures that did an inverse mapping between the users' direction of movement and the camera movement. For example, when they moved their finger to the left they expected the scene to move to the left, as if grabbing and dragging an image, which means that the camera will actually move to the opposite direction to achieve that effect. We called this metaphor world manipulation.

The mapping dimension describes the way in which the user's motion is mapped to a respective camera movement. We distinguished between position mapping, where the user actions have a direct effect on the position of the camera, and velocity mapping, where users increase or decrease the moving velocity of the camera through their actions. For example, if the user drags her finger by one cm to the left, in the first case, the camera will proportionally move by one meter to the left (or right depending on the navigation metaphor), and in the second case, it will continue moving to the left with a velocity proportional to the drag distance for as long as the user keeps her finger at the spot.

Finally, the gesture type follows the categories proposed in [32] to distinguish between those using the touch screen of the device, called surface gestures, and those using the accelerometers and rotation sensors, called motion gestures. There were some cases where a gesture combined motion while touching the screen. Then, the gesture was categorized as a motion gesture.

We categorized the proposed user gestures based on the aforementioned taxonomy. The results are displayed in Figure 6. Regarding the phone orientation, there seems to be a slight preference to portrait mode. In total, 58.9% of the gestures were performed in portrait mode, whilst the other 46.4% were in landscape. The most dominant case in the number of fingers was the use of a single finger, which appeared in 65.2% of the gestures compared to 30.4% using zero fingers and only 4.5% using two. As per the preferred angle, it is obvious from the results that the majority of users held the phone in a diagonal angle (52.7%), gestures performed with a horizontal phone setup were significantly fewer (42.9%), whilst only 4.5% of the gestures used a vertical angle. The dominant manipulation mode was viewpoint, which was the one performed in 82.1% of all gestures. However, there is a significant percentage of users who proposed gestures using a world metaphor; it was found in 17.9% of all gestures. Similarly, most gestures used velocity mapping (75.9%), but some users proposed actions with position mapping (found in 24.1% of all gestures). Finally, the type of gestures that the users performed was mostly surface gestures using single or multiple touches (67%). Motion gestures were significantly fewer (33%), and they were proposed mostly by inexperienced users.
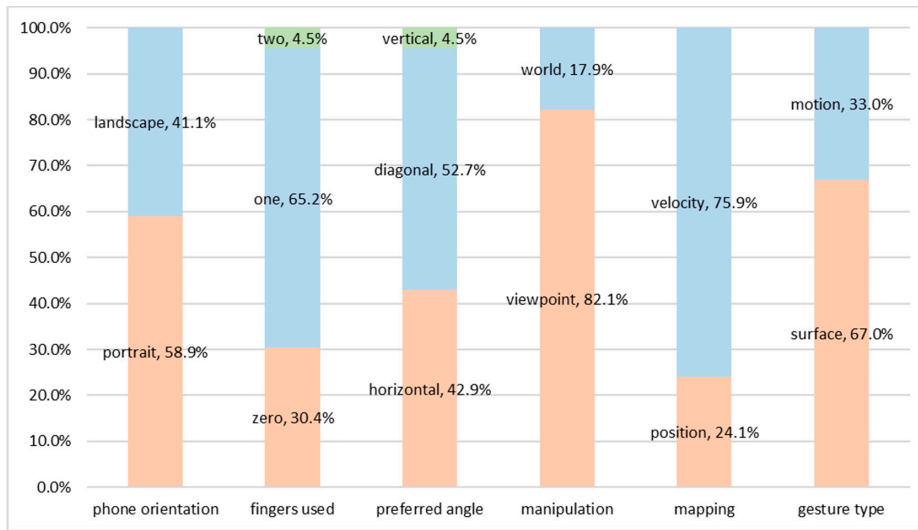
Figure 6. The taxonomy distribution of the elicited gestures.

3.2.2. Agreement Scores

Next, we examined all the gestures proposed by the users and identified the number of unique gestures per task. For the walk fwd-back task, we identified five unique gestures, for the rotate left-right task the unique gestures were seven, for look up-down, they were five, and for walk sideways, they were seven.

We calculated the agreement score, *AS*, for each task, *t*, based on the number of users that agreed for each of the unique gestures. We used the type proposed in [27]:

$$AS(t) \; = \; \Sigma_{i=1}^{n} \left(\frac{P_i(t)}{P(t)} \right)^2, \tag{1}$$

where *i* is the number of unique gestures, P_i is the number of users that agreed on the *i*th gesture, and *P* is the total number of users. The highest agreement score was reached in walk sideways (0.53), followed by walk fwd-back (0.43). On the contrary, the two remaining tasks associated with camera rotations, i.e., look up-down and rotate left-right, had lower agreement scores (0.36 and 0.24, respectively). The results are shown in Figure 7.
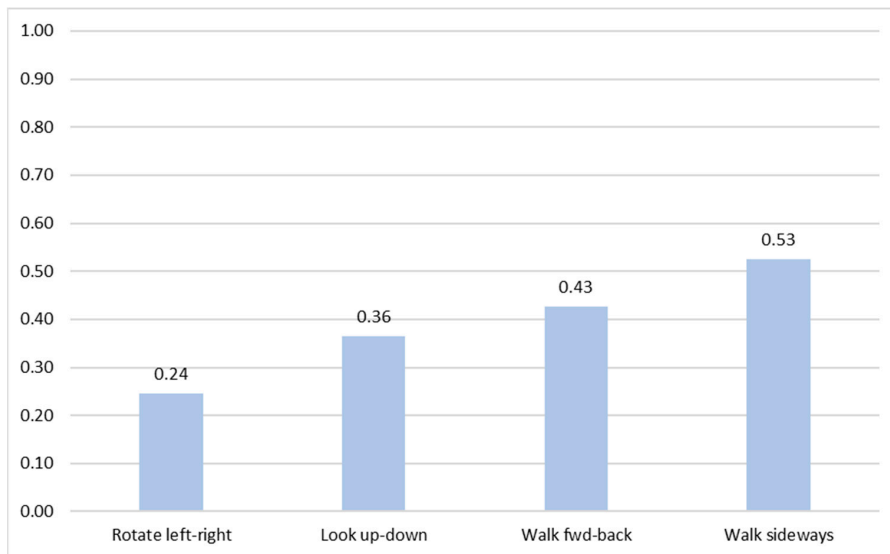

Figure 7. The agreement scores of the four navigation tasks arranged from lower to higher.

3.2.3. Subjective Ratings

Regarding the subjective user ratings of their proposed gestures, the scores were generally high. The mean values of the question 'how easy it was to produce the gesture' for the tasks, walk fwd-back, rotate left-right, look up-down, and walk sideways, were 4.82, 4.21, 4.32, and 4.39, respectively. We performed a Wilcoxon signed rank test between the easiness ratings of the four tasks and found that walk fwd-back was rated significantly higher compared to the other three tasks ($p = 0.004$ for rotate left-right, $p = 0.004$ for look up-down, and $p = 0.042$ for walk sideways). Regarding the appropriateness of the proposed gestures, people seemed to agree that their gestures were appropriate, but the mean score was noticeably lower for the look up-down gesture (3.75). The other three gestures had mean scores of 4.43 for walking fwd-back, 4.11 for rotating left-right, and 4.14 for walking sideways. A Wilcoxon signed rank test between the values of appropriateness of the four tasks revealed a statistically significant difference between look up-down and walk fwd-back ($p = 0.006$) and between look up-down and walk sideways ($p = 0.029$). Figure 8 displays the mean values of subjective user ratings for the four tasks with 95% confidence intervals.
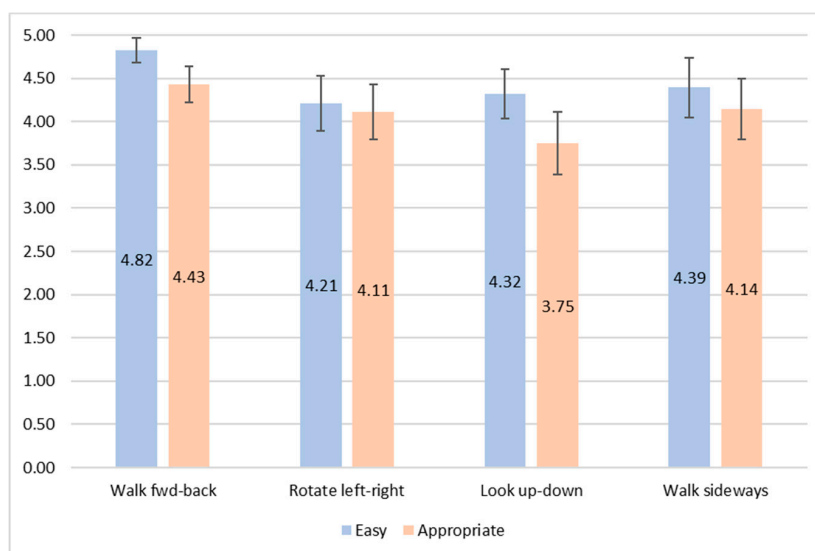

Figure 8. Mean values of subjective user ratings regarding the easiness and appropriateness of proposed gestures for each of the four tasks. Error bars represent 95% confidence intervals.

3.2.4. Candidate Gestures

The candidate gestures as proposed by the users in this study are summarized in Table 2. We considered a gesture as a candidate if more than half of the users agreed on it. Otherwise, we added the first two most agreed gestures as two alternative candidates. In all four tasks, the gesture with the most agreement was the 'virtual joystick' (walk fwd-back: 60.71%, rotate left-right: 39.29%, look up-down: 50%, and walk sideways: 71.43%). In these cases, users imagined dragging towards a direction and having the scene or the camera move or rotate towards this direction. A second alternative for both rotational tasks is to use motion gestures by turning the phone. For rotating to the left or right, users proposed to turn the phone to the left or right, respectively (horizontal tilt, 21.43%), and for looking up or down, to turn the phone upwards or downwards, respectively (vertical tilt, 32.14%).

Table 2. The candidate gestures for each of the four tasks.

Task	Candidate Gesture(s)	Description	Perc. of Agreement
Walk fwd-back	virtual joystick	Drag up or down to move forwards or backwards	60.71%
Rotate left-right	virtual joystick	Drag left or right to rotate	39.29%
	tilt horizontal	Rotate device to the left or right to turn	21.43%
Look up-down	virtual joystick	Drag up or down to look in the respective direction	50%
	tilt vertical	Rotate device upwards or downwards	32.14%
Walk sideways	virtual joystick	Drag left or right to move sideways	71.43%

3.2.5. Comments and Observations

We took note of some interesting user comments and observations regarding the use of the mobile phone during the study, which are reported below.

There were some users who attempted to explain their choice of phone orientation, which was related to their own background and expectations. Two users reported that the landscape mode seemed more natural/intuitive to them, because the projected environment is also displayed in landscape (widescreen). Others, mostly experienced gamers, instantly held the mobile phone in landscape mode and placed both their thumbs on the surface, because this grip and posture was to them like using a game controller. On the contrary, there was a user (F, 23) who preferred portrait mode because "it is not a gaming context, it is an installation in a public space and the interaction should not feel like playing a console game." She also added that we usually hold our phones in public with a single hand in portrait mode. A number of users were in favor of the portrait mode because "it is supposed to work as a remote control, and remote controls are held with a single hand." Finally, there was a user who reported that the process reminded him of controlling a drone through the phone, which was also reflected in his choice of the inverse motion direction for the look up and down gesture.

A few users explained their choice not to use any motion gestures by stating that from their previous experience with games or other control tasks, the interactions based on accelerometers were not as accurate as they expected, and they could not have fine control of the process. Thus, they preferred surface gestures although they admitted that they are less 'playful.'

The use of four degrees of freedom for navigation puzzled some users as they tried to fit multiple gestures and avoid conflicts. Actually, a significant percentage of the users reconsidered some of their previously proposed gestures when we moved on with new tasks during the study. This was more intense when they tested the combination of two tasks, i.e., walking while rotating (walk fwd-back combined with rotate left-right) and looking around (rotate left-right combined with look up-down). Some users proposed that there are two different rotate left-right gestures, one for walking and another for looking around. As one user put it there should be a gesture set for "controlling the body" and another one for "controlling the head." Furthermore, some users tried to fit the same gesture (e.g., virtual joystick or tilt) in different tasks using some means to switch between tasks. The proposed approaches for task-switching were: (a) To touch a button once to indicate the required mode before performing the gesture; (b) to keep touching a button/area (as a different mode compared to not pressing it); (c) to change phone orientation; and (d) to change automatically based on the context.

Finally, some users, especially less experienced ones, proposed some interesting and intuitive gestures. One user proposed to hold the phone horizontally and use it as a mini-map: To draw the path of the avatar on the surface using touch gestures. A few users held and used the phone as a pointing device to have the avatar look around the scene. Also, two users opted for a context-based interaction. They claimed they would expect the environment to "guide" them in order to minimize control actions, e.g., "when there is a painting I should look closer at, there could be an indicator that I can stand on a specific point and only then I will be able to look up and down." Figure 9 presents four pictures of the elicited gestures.

Figure 9. Four pictures of the elicited gestures. A user proposed two finger rotation for turning to the left or right (upper left); a typical gamer configuration with two virtual joysticks, one for each thumb (upper right); a user switched between the horizontal (lower left) and vertical (lower right) angle for walking and looking around, respectively.

4. Discussion

According to the results of the first study, it appears that navigating in 3D environments using a mobile phone performs at least as good as using body posture and mid-air gestures. It was considered as more comfortable, and it was also the preferred choice for most participants. This is an indication that mobile devices could be used as reliable alternatives to Kinect and similar depth-sensors for 3D navigation in public environments. This finding is in accordance to the results of the study presented in [7], where the smartphone is considered an attractive and stimulating solution for interactive applications in public settings. However, a further issue that needs to be investigated is whether visitors of museums and other public institutions are willing to download specific apps and use their personal phone as a controller.

The results of the elicitation study indicate a noticeable preference of surface over motion gestures when using the mobile phone as a navigation controller. Other usability studies have produced similar results, e.g., [29], but also studies where motion gestures have been rated better compared to multitouch, e.g., [33]. One possible explanation for our results is the fact that people are very well used to touch or multitouch actions in their daily experience with their mobile devices, e.g., for browsing content, interacting with interface elements, zooming and rotating media, etc., such that it is their first and most intuitive choice. Furthermore, as mentioned before, some more experienced users are affected by previous experiences with tilt-based control and they prefer to avoid it as it is less accurate.

Another interesting result of the study was a certain clustering between experienced gamers and non-gamers that was detected. The majority of people from the first group tried more or less to replicate the functionality of console game controllers or of 3D mobile game interfaces, holding the phone in a landscape orientation and proposing two virtual joysticks, one for each thumb. On the contrary, less experienced users were keener on using motion gestures, they used their imagination more, and in some cases, they proposed really intuitive solutions—although not all of them were feasible. Some of them proposed to have more automated actions and assistance, even if it means sacrificing some control or degrees of freedom. This finding is consistent with the results of previous studies [34] that indicate the fact that first-person walking can be confusing for inexperienced users, raising a need for effective navigation aids. Interactive systems of this kind should preferably be designed to monitor user performance and, in case they detect non-expected behavior, to offer assistance in a non-intrusive way, e.g., through messages, indications, mini-maps, etc.

In some cases, users tried to bring familiar 2D gestures to some of the navigation tasks, resulting in the world manipulation metaphor (also termed "scene in hand") mentioned before. For example, for walking sideways, they imagined dragging on the phone to the direction where the image should be moving, as if they are performing a horizontal pan action. Similarly, they proposed dragging with their finger down on the phone to move forward, as if they are grabbing the ground floor and moving it towards them. For some of these cases, we tried again in the opposite direction through the Wizard of Oz technique, but the users reported that it was confusing for them. A possible reason that some users chose a 2D gesture is that they had to perform single-axis tasks one by one during the study. So, inexperienced users who did not foresee the rest of the actions needed for navigation in 3D environments tended to assign them to familiar 2D interactions.

Finally, we have to mention that the results of this study are possibly applicable to wearable VR environments as well, as it has been already showcased that a second mobile device can be effectively used as a controller for a cardboard-based virtual museum [35]. In that case, there is, however, a couple of important differences regarding the user interactions and context of use. First, two degrees of freedom (look up-down and rotate left-right) can be directly controlled through the user's head rotation, so there is no need to include them in the mobile controller. The controller could therefore include extra functions for selecting or manipulating content. Second, given that the users will not be able to directly look at the screen of their device due to the headset, the interactions must be carefully designed, probably also using appropriate feedback-sound or haptics-, so that they can be performed 'blindly.'

5. Conclusions

This paper examined the suitability of mobile devices as controllers for first-person navigation in 3D by presenting a comparative study between Kinect-based and smartphone-based interactions, followed by a gesture elicitation study for suitable navigation gestures using a smartphone. The results of the first study indicated that the smartphone-based input is at least as reliable as the Kinect-based input. It was preferred by most participants, considered more comfortable, rated higher in all subjective ratings, and produced significantly shorter task completion times in the first of the two scenarios. Furthermore, the second study highlighted the preferred input techniques for 3D navigation using a mobile device, which were considerably different from the techniques we designed for our first study. It also produced some interesting observations regarding the different expectations of users and their preferred ways of mapping their actions to 3D movements, based on their previous experience and background.

An important advantage of a modern smartphone compared to typical game controllers is that its interface can be fully customized, e.g., by adding custom virtual buttons on the screen or by supporting various multitouch gestures. As such, it is more adaptive to afford multiple interaction modalities with a 3D application. Designers could consider creating two or more alternative setups to support various user groups, e.g., experienced vs. non-experienced, with various levels of control and support, and even different gesture types.

An issue that needs further research regarding the design of 3D navigation techniques is the support of secondary actions. Often navigation has to be combined with other actions, such as selecting an object, browsing information, etc. Therefore, the interaction techniques designed for 3D navigation should leave room for other tasks that can be performed in parallel. This is an aspect that was not considered in our studies.

In the future, we are planning to further explore the prospects of combining smartphone-based interactions with virtual environments in public settings. We aim to design and develop a number of alternative interfaces based on the results of our elicitation study and compare them in terms of usability and efficiency. Furthermore, we are planning to extend the functionality of the proposed controller by adding non-intrusive user assistance through voice messages and icons, and by including additional forms of feedback, e.g., audio and vibration.

Appendix A

Questions for subjective user ratings used in the comparative study:
[for each input modality]

1. (ease-of-use) I thought the technique was easy to use [1-strongly disagree to 5-strongly agree]
2. (learnability) I learned the technique quickly [1-strongly disagree to 5-strongly agree]
3. (satisfaction) I was satisfied with the use of the technique [1-strongly disagree to 5-strongly agree]
4. (comfort) I felt comfortable using the technique [1-strongly disagree to 5-strongly agree]
5. (accuracy) I could navigate with great accuracy using the technique [1-strongly disagree to 5-strongly agree]

Questions for subjective user ratings used in the gesture elicitation study:
[for each gesture]

1. (easy) How easy was it for you to produce this gesture [1-very difficult to 5-very easy]
2. (appropriate) How appropriate is your gesture for the task [1-not appropriate at all to 5-very appropriate]

References

1. Kostakos, V.; Ojala, T. Public displays invade urban spaces. *IEEE Pervasive Comput.* **2013**, *12*, 8–13. [CrossRef]
2. Hornecker, E.; Stifter, M. Learning from interactive museum installations about interaction design for public settings. In Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments, Sydney, Australia, 20–24 November 2006; pp. 135–142.
3. Castro, B.P.; Velho, L.; Kosminsky, D. INTEGRARTE: Digital art using body interaction. In Proceedings of the Eighth Annual Symposium on Computational Aesthetics in Graphics, Visualization, and Imaging, Annecy, France, 4–6 June 2012.
4. Häkkilä, J.; Koskenranta, O.; Posti, M.; He, Y. City Landmark as an Interactive Installation: Experiences with Stone, Water and Public Space. In Proceedings of the 8th International Conference on Tangible, Embedded and Embodied Interaction, Munich, Germany, 16–19 February 2014; pp. 221–224.
5. Grammenos, D.; Drossis, G.; Zabulis, X. Public Systems Supporting Noninstrumented Body-Based Interaction. In *Playful User Interfaces*; Springer: Singapore, 2014; pp. 25–45.
6. Krekhov, A.; Emmerich, K.; Babinski, M. Gestures from the Point of View of an Audience: Toward Anticipatable Interaction of Presenters with 3D Content. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; pp. 5284–5294.
7. Bergé, L.P.; Perelman, G.; Raynal, M.; Sanza, C.; Serrano, M.; Houry-Panchetti, M.; Cabanac, R.; Dubois, E. Smartphone-Based 3D Navigation Technique for Use in a Museum Exhibit. In Proceedings of the Seventh International Conference on Advances in Computer-Human Interactions (ACHI 2014), Barcelona, Spain, 23–27 March 2014; pp. 252–257.
8. Cristie, V.; Berger, M. Game Engines for Urban Exploration: Bridging Science Narrative for Broader Participants. In *Playable Cities*; Springer: Singapore, 2017; pp. 87–107.
9. Darken, R.P.; Sibert, J.L. Wayfinding strategies and behaviors in large virtual worlds. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 13–18 April 1996; Volume 96, pp. 142–149.
10. Bowman, D.A.; Koller, D.; Hodges, L.F. Travel in Immersive Virtual Environments: An Evaluation of Viewpoint Motion Control Techniques Georgia Institute of Technology. In Proceedings of the 1997 Virtual Reality Annual International Symposium, Washington, DC, USA, 1–5 March 1997; pp. 45–52.

11. Vosinakis, S.; Xenakis, I. A Virtual World Installation in an Art Exhibition: Providing a Shared Interaction Space for Local and Remote Visitors. In Proceedings of the Re-thinking Technology in Museums, Limerick, Ireland, 26–27 May 2011.

12. Rufa, C.; Pietroni, E.; Pagano, A. The Etruscanning project: Gesture-based interaction and user experience in the virtual reconstruction of the Regolini-Galassi tomb. In Proceedings of the 2013 Digital Heritage International Congress, Marseille, France, 28 October–1 November 2013; pp. 653–660.

13. Cho, N.; Shin, D.; Lee, D.; Kim, K.; Park, J.; Koo, M.; Kim, J. Intuitional 3D Museum Navigation System Using Kinect. In *Information Technology Convergence*; Springer: Dordrecht, The Netherlands, 2013; pp. 587–596.

14. Roupé, M.; Bosch-Sijtsema, P.; Johansson, M. Interactive navigation interface for Virtual Reality using the human body. *Comput. Environ. Urban Syst.* **2014**, *43*, 42–50.

15. Hernandez-Ibanez, L.A.; Barneche-Naya, V.; Mihura-Lopez, R. A comparative study of walkthrough paradigms for virtual environments using kinect based natural interaction. In Proceedings of the 22nd International Conference on Virtual System & Multimedia (VSMM), Kuala Lumpur, Malaysia, 17–21 October 2016; pp. 1–7.

16. Guy, E.; Punpongsanon, P.; Iwai, D.; Sato, K.; Boubekeur, T. LazyNav: 3D ground navigation with non-critical body parts. In Proceedings of the 3DUI 2015, IEEE Symposium on 3D User Interfaces, Arles, France, 23–24 March 2015; pp. 43–50.

17. Ren, G.; Li, C.; O'Neill, E.; Willis, P. 3D freehand gestural navigation for interactive public displays. *IEEE Comput. Graph. Appl.* **2013**, *33*, 47–55. [PubMed]

18. Dias, P.; Parracho, J.; Cardoso, J.; Ferreira, B.Q.; Ferreira, C.; Santos, B.S. Developing and evaluating two gestural-based virtual environment navigation methods for large displays. In *Lecture Notes in Computer Science*; Springer: Singapore, 2015; Volume 9189, pp. 141–151.

19. Kurdyukova, E.; Obaid, M.; André, E. Direct, bodily or mobile interaction? Comparing interaction techniques for personalized public displays. In Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia (MUM'12), Ulm, Germany, 4–6 December 2012.

20. Baldauf, M.; Adegeye, F.; Alt, F.; Harms, J. Your Browser is the Controller: Advanced Web-based Smartphone Remote Controls for Public Screens. In Proceedings of the 5th ACM International Symposium on Pervasive Displays, Oulu, Finland, 20–26 June 2016; pp. 175–181.

21. Medić, S.; Pavlović, N. Mobile technologies in museum exhibitions. *Turizam* **2014**, *18*, 166–174. [CrossRef]

22. Vajk, T.; Coulton, P.; Bamford, W.; Edwards, R. Using a mobile phone as a "Wii-like" controller for playing games on a large public display. *Int. J. Comput. Games Technol.* **2018**, *2008*. [CrossRef]

23. Du, Y.; Ren, H.; Pan, G.; Li, S. Tilt & Touch: Mobile Phone for 3D Interaction. In Proceedings of the 13th international conference on Ubiquitous computing, Beijing, China, 17–21 September 2011; pp. 485–486.

24. Shirazi, A.S.; Winkler, C.; Schmidt, A. Flashlight interaction: A study on mobile phone interaction techniques with large displays. In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09), Bonn, Germany, 15–18 September 2009.

25. Joselli, M.; Da Silva, J.R.; Zamith, M.; Clua, E.; Pelegrino, M.; Mendonça, E.; Soluri, E. An architecture for game interaction using mobile. In Proceedings of the 4th International IEEE Consumer Electronic Society—Games Innovation Conference (IGiC 2012), Rochester, NY, USA, 7–9 September 2012; pp. 1–5.

26. Liang, H.; Shi, Y.; Lu, F.; Yang, J.; Papangelis, K. VRMController: An input device for navigation activities in virtual reality environments. In Proceedings of the 15th ACM SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry, Zhuhai, China, 3–4 December 2016; pp. 455–460.

27. Wobbrock, J.O.; Aung, H.H.; Rothrock, B.; Myers, B.A. Maximizing the guessability of symbolic input. In Proceedings of the CHI '05 Extended Abstracts on Human Factors in Computing Systems, Portland, OR, USA, 2–7 April 2005; pp. 1869–1872.

28. Du, G.; Degbelo, A.; Kray, C.; Painho, M. Gestural interaction with 3D objects shown on public displays: An elicitation study. *Int. Des. Archit. J.* **2018**, *38*, 184–202.

29. Obaid, M.; Kistler, F.; Häring, M.; Bühling, R.; André, E. A Framework for User-Defined Body Gestures to Control a Humanoid Robot. *Int. J. Soc. Robot.* **2014**, *6*, 383–396. [CrossRef]

30. Wobbrock, J.O.; Morris, M.R.; Wilson, A.D. User-defined gestures for surface computing Proceedings of the 27th international conference on Human factors in computing systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'09), Boston, MA, USA, 4–9 April 2009; p. 1083.

31. LaViola, J.J., Jr.; Kruijff, E.; McMahan, R.P.; Bowman, D.; Poupyrev, I.P. *3D User Interfaces: Theory and Practice*; Addison-Wesley Professional: Boston, MA, USA, 2017.

32. Ruiz, J.; Li, Y.; Lank, E. User-defined motion gestures for mobile interaction. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 197–206.

33. Pazmino, P.J.; Lyons, L. An exploratory study of input modalities for mobile devices used with museum exhibits. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (CHI'11), Vancouver, BC, Canada, 7–12 May 2011; pp. 895–904.

34. Burigat, S.; Chittaro, L. Navigation in 3D virtual environments: Effects of user experience and location-pointing navigation aids. *Int. J. Hum. Comput. Stud.* **2007**, *65*, 945–958. [CrossRef]

35. Papaefthymiou, M.; Plelis, K.; Mavromatis, D.; Papagiannakis, G. *Mobile Virtual Reality Featuring A Six Degrees of Freedom Interaction Paradigm in A Virtual Museum Application*; Technical Report; FORTH-ICS/TR-462; FORTH: Crete, Greece, 2015.