Exploring Smartphone-Based Interaction with Overview+Detail Interfaces on 3D Public Displays

Louis-Pierre Bergé, Marcos Serrano, Gary Perelman, Emmanuel Dubois
IRIT – University of Toulouse
118 Route de Narbonne, 31062 Toulouse, France
{Louis-Pierre.Berge, Marcos.Serrano, Gary.Perelman, Emmanuel.Dubois}@irit.fr

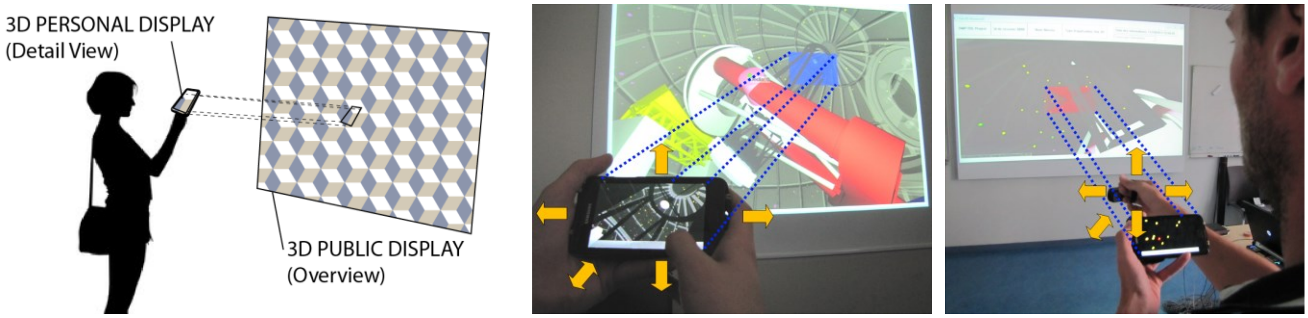


Figure 1: a) General setting of smartphone-based Overview+Detail interface on a 3D Public Display. We used two mid-air navigation techniques in a public installation to explore a 3D telescope visualization: b) Mid-Air Phone and c) Mid-Air Hand.

ABSTRACT

As public displays integrate 3D content, Overview+Detail (O+D) interfaces on mobile devices will allow for a personal 3D exploration of the public display. In this paper we study the properties of mobile-based interaction with O+D interfaces on 3D public displays. We evaluate three types of existing interaction techniques for the 3D translation of the Detail view: touchscreen input, mid-air movement of the mobile device (Mid-Air Phone) and mid-air movement of the hand around the device (Mid-Air Hand). In a first experiment, we compare the performance and user preference of these three types of techniques with previous training. In a second experiment, we study how well the two mid-air techniques perform with no training or human help to imitate usual conditions in public context. Results reveal that Mid-Air Phone and Hand perform best with training. However, without training or human help Mid-Air Phone is more intuitive and performs better on the first trial. Interestingly, on both experiments users preferred Mid-Air Hand. We conclude with a discussion on the use of mobile devices to interact with public O+D interfaces.

Author Keywords

Overview + detail; 3D interfaces; public display; mid-air interaction; personal displays

ACM Classification Keywords

H.5.2. Information interfaces and presentation: Interaction.

MobileHCI '14, September 23 - 26 2014, Toronto, ON, Canada.

http://dx.doi.org/10.1145/2628363.2628374

INTRODUCTION

Public displays allow pedestrians to interact with 2D content such as city maps or tourism information [29]. In this context, existing systems have used smartphones as personal views (Detail) of the public display (Overview), leveraging multi-user access to one display [12]. Visualisation of 3D content on public displays is emerging to visualize scientific 3D data [37]; to explore culture heritage 3D scanned objects [1]; to teach history through virtual tours of 3D reconstructed historical sites [1]; to play public 3D games [42] or to navigate a city 3D map [44].

Most of these examples already include the use of a personal device to interact with the 3D content (to orient and position a slice plane [37] or to navigate in the 3D environment [44]) but few consider how to apply the Overview and Detail (O+D) paradigm using the smartphone. Using the O+D on mobile devices will provide the user with the ability to privately visualize details of the 3D environment while still taking advantage of the mobile device input to interact with the 3D content. However, the public context imposes certain constraints in terms of user's profiles (mainly beginners) and appropriate interaction techniques (which need to be easy to understand and perform).

In this paper we focus on the translation task, i.e. how to move the Detail view (displayed on the smartphone) on a 3D environment, the Overview (displayed on the public display). We report on the evaluation of three interaction techniques for controlling the translation of the Detail view: mid-air hand gestures around the mobile device like [27], mid-air movements of the mobile device itself and classical touchscreen pad. We performed two experiments. The first one is a controlled experiment with training: its goal is to evaluate the three techniques in terms of performance, usability and user preference. The second experiment aims at studying the difficulty of performing the two mid-air gesture-based techniques in a situation imitating a public space, i.e. without the experimenter help or training.

Our contributions include: 1) An elicitation of the design space for mobile 3D O+D, 2) a controlled evaluation of three techniques for the 3D translation of the Detail view and 3) an evaluation of the difficulty to perform the mid-air techniques with no human help or training. We also report about an exhibition in public context for several days.

RELATED WORK

Our research is inspired by previous works on using mobile devices to visualize large datasets, to navigate 3D scenes and to interact with public displays.

Large information space visualization on mobile phone

To explore large information spaces such as web pages or maps, research on visualization proposes three solutions to separate focused and contextual views: Overview+Detail (spatial separation), Zooming (temporal separation) and Focus+Context (seamless focus in context) [12]. On desktop environments, previous research established that Overview+Detail (O+D) is worse than Zooming or Focus+Context (F+C) in terms of task completion time [4, 12]. However, users usually prefer O+D because it helps them build a mental model of the explored information [31]. In addition, display size has less impact on O+D than Zooming or F+C [34].

Recent implementations of these types of interfaces involved mobile devices [10] to support information search on a webpage [35], a map [9] or a scatterplot [40]. Results are mixed: O+D is faster with off-screen objects and maps but slower with scatterplots. And yet, participants prefer O+D for navigation on webpages [35]. They also are better in spatial recall of targets when using O+D [9]. Still, none of these previous works explored the use of focused or contextual views for 3D content.

Our goal is to explore interaction with a large screen providing the 3D contextual view and a mobile phone displaying the 3D focused view. F+C and Zooming interface are not in line with our setting because we have two separate views. Moreover, using F+C interface involve intentional distortion view such as fisheye or lens [12]. Using Zooming interface raises the risk of perturbing user's position and losing the user in the 3D volume. O+D interface seems to be good alternative for our context.

3D navigation with a mobile phone

We focus our analysis on navigation techniques for 3D content [8] on smartphones and where the user controlled the navigation, i.e. active navigation. Existing solutions can be divided into three main types: using the touchscreen, moving the smartphone and interacting around the smartphone (around-device interaction).

Touchscreen is the standard input on mobile 3D games, usually through a rate-based pad [39]. Research works explored touchscreen input with one [5, 13, 18] or two fingers [10, 39]. To reduce finger occlusion, Hachet et al. extended the mobile phone with a 3-DOF elastic controller attached to the device [17].

Using built-in sensors or adding new ones, researchers proposed different techniques based on moving the smartphone. Hurst et al. explored two different metaphors [22]: fixed world, where the 3D environment does not move in relation with the real world, and "shoebox", where the 3D environment is fixed to the smartphone. The idea of pointing with the smartphone as if it was a magic lens has been used for augmented reality applications [1, 32] or to interact with virtual volumetric data [38]. A less direct solution based on tilting gestures has been compared with touchscreen input to navigate a virtual reality panoramic [21]. Results show that the gesture-based technique performed twice as well as the touchscreen input in an orientation task and was preferred by 80% of the users. More recent works investigated around-device interaction as a novel approach to interact with 3D content [36, 26, 27]. PalmSpace is a mid-air gesture technique to rotate 3D objects on the mobile display [27]. Around-device gestures have also been used to navigate 2D multi-scale maps [26].

In our research we compare these three different types of interactions in the context of 3D O+D public displays.

Public Display and mobile phone

To interact with public displays, researchers explored using multi-touch input [14] or mid-air gestures [43]. Another strategy inspired by the prevalence of mobile devices is the use of smartphones as remote controllers [28, 30]. For instance, touchscreen input is used to control a distant cursor [28], to select an item [7] or to pan and zoom [30]. Mid-air gesture with the phone has also been used for some of these tasks [6, 32].

Few works explored interaction with 3D content on public displays. Smartphones have been used to reveal hidden 3D content on augmented posters [16]. In the context of large displays (not necessarily in public context), Song et al. investigated the use of multi-touch interaction and mid-air gestures to explore and annotate 3D data [37].

An important aspect to consider on public displays is how to reveal the interaction technique to the user. This aspect has been barely addressed in the state of the art. Walter et al. [43] investigated different strategies to reveal mid-air gestures to interact with a public display. Their study indicates that users intuitively discover gestures by imitating or extending other user's gestures.

Our work is inspired by solutions using the mobile phone as the means to interact with public displays. For the task of interacting with a 3D O+D interface, existing interaction approaches need to be compared and evaluated. In particular, the public context imposes constraints on how to reveal and learn these interaction techniques. The goal of our work is to deepen existing research on using mobile phones to interact with 3D interfaces on public displays.

INTERACTION TECHNIQUE DESIGN: APPROACH

Among the considerations addressed in the literature, we focus on the main properties for mobile-based interaction with 3D Overview+Detail (O+D) interfaces.

3D Tasks

In our work we focus on the 3D navigation task and more precisely the *travel* subcategory defined by Bowman [8]. Generally the user controls 3 degrees of freedom (DOF) to translate the point of view and 3-DOF to rotate the point of view. Many metaphors have been explored to reduce these DOFs, such as the flying vehicle control metaphor or avatar metaphor [8]. In our work, we limit user control of the Detail view to a 3-DOF translation. This task is sufficient to explore public 3D content such as museum objects and it simplifies the task in a public setting, where interaction needs to be intuitive and straightforward.

Types of interaction techniques

Works presented in the literature report about different approaches to interact with 3D content. Among them using a touchscreen [18], moving the device [5] or around the device [27] have been implemented on mobile phone. We explore these three different approaches to control the position of the Detail view:

- **Mid-Air Hand** (Figure 2–a): the position of smartphone serves as a spatial reference. The position of the hand in this referential is mapped to the virtual position of the Detail view. We constrain the movement of the hand to the area behind the mobile phone. A virtual button on the mobile screen (de)activates this navigation mode.
- **Mid-Air Phone** (Figure 2–b): similar to [22], translations applied to the mobile phone translate the virtual position of the Detail view. As for Mid-Air Hand, a virtual button on the mobile display (de)activates this navigation mode.
- **Touchscreen** (Figure 2–c): inspired by commercial mobile 3D games, we use two rate-based joysticks to control the virtual position of the Detail view. The left circular joystick controls the 2D translation along the X and Y axis. The right cylinder joystick controls the 1D translation along the Z axis. Both pads can be used at the same time to control the 3-DOF navigation.

Mapping

Two ways of mapping a gesture with a 3D translation are illustrated in the literature: direct or indirect [26, 38]. In the direct mapping, the absolute position or gesture offset (in mid-air) is directly mapped to a 3D position on the virtual environment. Both the Mid-air Phone and the Mid-air Hand techniques are based on direct mapping. In the indirect mapping, the gesture controls the velocity and direction of the camera movement in the virtual environment. The Touchscreen technique is based on indirect mapping.

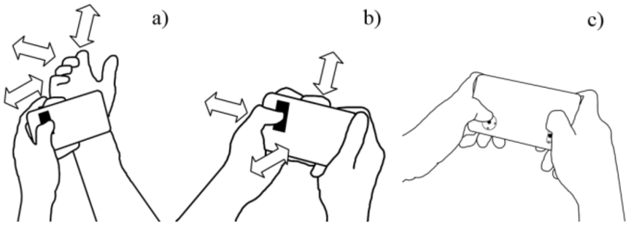


Figure 2: Three types of techniques: a) Mid-Air Hand, b) Mid-Air Phone and c) Touchscreen.

Mid-air gestures reference

Another factor concerning mid-air gestures is whether these gestures relate to a reference position in the real world. We can divide gestures into absolute or relative. Absolute gestures are not linked to a particular position in the real world. For instance, the Mid-Air Phone is absolute: when the user presses the button the virtual camera moves according to the absolute displacement of the phone in the real world. Relative mid-air gestures are linked to a real-world reference object. The Mid-Air Hand technique is relative as it uses the phone as a reference: the movements of the hand are linked to the position of the phone and limited in space around this reference. Previous work suggests that using a reference is useful for mid-air input without visual feedback [15] and for 3D interaction [20].

O+D visualization

To design the representational aspect of O+D interfaces, a set of parameters can be adjusted [10]: 1) the overlap of O+D views (usually not overlapping), 2) the relative size of both views (Overview is usually smaller than Detail view, as in PowerPoint or Google Maps), 3) the volume of space displayed by the Detail view, 4) the number of Detail views and 5) the feedback of the Detail view's position in the Overview (typically a polygonal outline or a shaded area).

In our work, we fixed these visual parameters in order to focus on input interaction. Thereby, the two views are non-overlapping and are displayed on two different displays: the Overview on the Public Display and the Detail view on the smartphone. As opposed to usual O+D, the Detail view is smaller than the Overview. In public context, this setting allows a personal view and multi-user access to the public display. The Detail view displays the 3D environment with no depth limitation. In terms of feedback, a coloured 3D pyramid on the Overview indicates the position of the Detail view: one color is assigned to each user.

Revealing message

Another important factor on public displays is how to inform the user of the interaction to perform. This has already been explored for full-body interaction with a large public display [43]. Different types of feedback (text, icon, video…) were explored. In our work we designed textual and image-based revealing messages to assess the suitability of studied interactions to public context.

3D environment: occlusion and orientation

Previous work on the use of O+D interfaces mainly refers to large 2D information spaces. In a 3D environment, interaction with the Detail view can be difficult due to occlusion and orientation issues. 3D elements can occlude the target or the feedback of the Detail view in the Overview. Without feedback, it may be hard or almost impossible for the user to situate the Detail view position in the 3D environment. To study this factor, our first experiment includes two conditions: with and without occlusion objects.

In next sections we report on two experiments to explore these properties. The first study evaluates the three types of gestures with previous training and with different occlusions. The second study evaluates the two mid-air techniques without training or indications from the experimenter.

SYSTEM DEVELOPMENT

In this section we describe the implementation of the 3D O+D environment and the three interaction techniques illustrated in Figure 2.

3D environment

Our implementation runs synced 3D content both on a mobile device and on a large display. We used a Samsung Galaxy S2 smartphone (6.6x12.5x0.8cm, 116 gr., 4.3” screen) running Android 4.1.2. The large display was a 24” monitor with a resolution of 1920x1080px. To implement all 3D content, we used a C++ open source engine based on OpenGL, Irrlicht [23]. The same C++ 3D code executed on the smartphone and on the large display: to run it on Android we compiled it with a JNI-based tool. The 3D scene on the smartphone ran at 20fps and the scene on the large display ran at 300fps. This difference of frequency was barely perceivable and no user commented on this.

Figure 3. Tracker on hand and on smartphone (left) and user performing the Mid-Air Hand technique (right).

Mid-air gestures tracking

To prototype the mid-air gestures we used two Polhemus Patriot Wireless trackers (7x3x2.5cm, 79.4gr.) [33]. One tracker was attached to the back of the smartphone using Velcro touch fastener. For the Mid-Air Hand technique, the tracker was attached to the hand using a glove and Velcro (Figure 3). We consider in the discussion section envisioned solutions to implement the mid-air tracking.

The trackers communicate wirelessly with a magnetic receptor, USB-connected to the computer running the 3D scene on the large display. The computer and the mobile phone communicated wireless through sockets using the IVY-bus library [24]. We filtered data using the 1€ filter [11]. The tracking system runs at a maximum frequency of 50Hz. Overall, the latency was negligible and no user commented on this.

EXPERIMENT 1: TRANSLATION OF THE DETAIL VIEW

The goal of this experiment is to evaluate the comparative performance of the three techniques presented in the previous section for the 3D translation of the Detail view.

Task and mapping

We asked users to reach a target on a 3D scene using the Detail view. A sphere represented the target. Virtual walls limited the 3D scene to a cubic area. We divided the cube into a matrix of 2x9 equal smaller cubes. We placed the target in the center of one of the 18 small cubes (Figure 4-left). The Detail view, displayed on the mobile screen, is initially situated on the front-center position (red cube in Figure 4-left). The Detail view does not show the entire 3D scene and so the user needs to look at the large display to find the target.

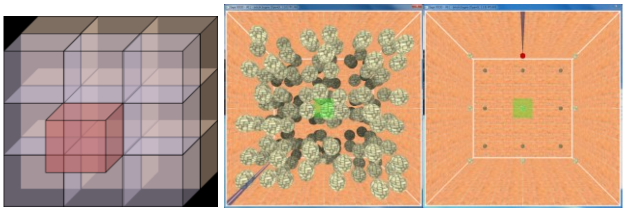


Figure 4: The 3D environment was divided into 18 areas (left), and displayed with or without occlusion objects (center/right).

The participant moves the Detail view until it is close enough to the target: on the Detail view, the target color changes to indicate the user reached it. We studied two occlusion conditions: with occlusion objects (10 spheres randomly distributed inside each of the 18 smaller areas) and without occlusion objects (Figure 4-center, right). We added a feedback on the target position: a ray attached to the target, parallel to the depth axis of the square area.

Apparatus

We used the apparatus described in the previous section.

Participants

We recruited 12 participants (3 female) aged 28.6 years on average (SD=5.2). Two of them had previously played 3D games on mobile phones and 10 of them had used mid-air interaction (Wiimote or Kinect).

Design and procedure

The experiment followed a 3x2 within-participant design, with *Technique* (Mid-Air Hand, Mid-Air Phone or Touchscreen pad) and *Occlusion* (True or False) as factors. Three blocks were run for each technique, the Techniques factor being counterbalanced by means of a 3X3 Latin Square. Trials in a block were grouped by the Occlusion factor, which was always ordered by increasing difficulty (first without Occlusion, then with Occlusion). Each block of trials required 18 selections per Occlusion factor. The 18 targets were randomly ordered. Each subject performed 3 techniques x 3 blocks x 2 occlusions x 18 targets = 324 trials. On average, the experiment lasted 97 minutes, with 1 min. 21 sec. of training for each *Technique*. The training period was performed before each technique and consisted of a minimum of 5 successful trials.

Collected data

We logged all tracking data as well as touch events from the touchscreen. We measured trial completion time from stimulus onset to target reached. Beside time, we collected user preference through a set of techniques: usability via the System Usability Scale questionnaire (SUS) [3] and attractiveness via the AttrakDiff questionnaire [19]. AttrakDiff [19] informs on the attractiveness of a technique according to three distinct dimensions: the pragmatic quality (PQ) indicates whether the user can achieve his goals; the hedonic quality (HQ) indicates to what extent the technique enhances the possibilities of the user; finally the attractiveness (ATT) gives an idea on how the user values each technique based on its quality and engagement.

EXPERIMENT 1: RESULTS

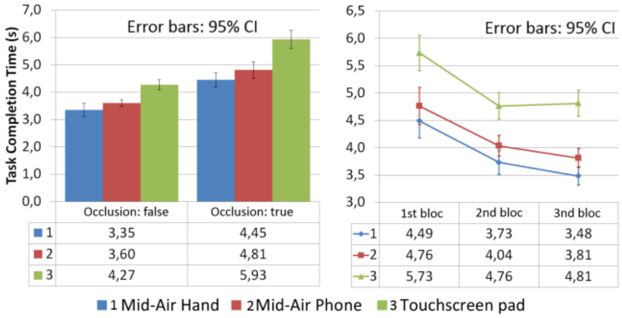
We collected 324 trials/user x 12 users = 3888 trials in total.

Quantitative results

Task completion times

A Shapiro-Wilk test shows that task completion time data does not follow a normal distribution ($p < 2.2 \times 10^{-16}$). We did not find any data transformation that would allow us to use parametric tests. Our statistical analysis is thus based on non-parametric tests. Figure 5 (left) summarizes the task completion times for each technique with and without occlusion. Three different Wilcoxon tests reveal that the occlusion factor has a significant effect on task completion time for each interaction technique: +32.8% (+1.1s, p<0.01) for Mid-Air Hand, +33.6% (+1.21s, p<0.01) for Mid-Air Phone and +38.9% (+1.66s, p<0.01). As expected, the completion time is longer with occlusion.

For the condition without occlusion, a Friedman test reveals a significant effect of the interaction technique on task completion times ($\chi^2(2)$=7.17, p=0.028). A post-hoc test using Wilcoxon with Bonferroni correction shows a significant difference between Mid-Air Hand and Touchscreen Pad (p=0.032). Without occlusion, only one direct mapping technique is significantly faster than the indirect mapping technique. For the condition with occlusion, a Friedman test reveals a significant effect of the interaction technique ($\chi^2(2)$=15.17, p<0.01). A post-hoc test using Wilcoxon with Bonferroni correction shows a significant difference between Mid-Air Hand and Touchscreen Pad (p<0.01) and between Mid-Air Phone and Touchscreen Pad (p=0.016).



	Occlusion: false	Occlusion: true
🟦 1	3,35	4,45
🟥 2	3,60	4,81
🟩 3	4,27	5,93

	1st bloc	2nd bloc	3nd bloc
1	4,49	3,73	3,48
2	4,76	4,04	3,81
3	5,73	4,76	4,81

Figure 5: Task completion times (in s) for each technique according to the Occlusion factor (left) and learning effect for each technique (right).

Learning effect

Friedman tests reveal a significant effect of the block order on task completion time for each different interaction technique (Mid-Air Hand: $\chi^2(2)$=10.67, p<0.01; Mid-Air Phone: $\chi^2(2)$=8.17, p=0.017; Touchscreen: $\chi^2(2)$=18, p<0.01). A post-hoc test using Wilcoxon with Bonferroni correction showed a significant difference between the first and the last block for all techniques (Figure 5-right). This confirms a learning effect. Completion time improves 22.5% for Mid-Air Hand (p=0.003), 20.0% for Mid-Air Phone (p=0.05) and 16.0% for Touchscreen (p<0.0001). Further studies are required to establish if a longer use of these techniques would increase the observed improvement as suggested in Figure 5 (right), especially for the Mid-Air Hand and Mid-Air Phone techniques.

Qualitative results

Three aspects have been considered in the qualitative evaluation: usability, attractiveness and user preference.

Usability evaluation

A SUS score was computed for each technique [3]: 81.04 (SD=12.94) for the Mid-Air Hand, 77.50 (SD=18.59) for the Mid-Air Phone and 72.5 (SD=22.69) for the Touchscreen. A Friedman test did not reveal any significant effect of the interaction techniques on the SUS score ($\chi^2(2)$=0.13, p=0.94). According to [3] the usability of the three techniques can be rate as “good”.

Attractiveness

To measure the attractiveness of the three techniques we relied on the Attrakdiff method [19]. We summarize in Figure 6-a, the results of the Pragmatic Quality (PQ) and Hedonic Quality (HQ) dimensions. According to the Attrakdiff report, Mid-Air Hand is rated as “desired”. More precisely, with regards to PQ, the technique is very pragmatic and assists the user optimally. With regards to HQ the report establishes the technique is very hedonic: the user identifies with the technique, which motivates and stimulates him. On the other hand, Mid-Air Phone is rated as “task-oriented”. With regards to PQ and HQ dimensions, the Attrakdiff report concludes there is room for improvement in terms of usability and user’s stimulation. Finally the Touchscreen technique is rated as “neutral”:

there is also room for improvement in terms of usability and stimulation.

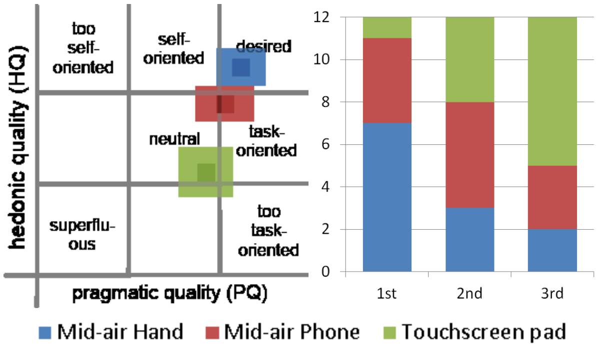


Figure 6: a) Portfolio generated using the AttrakDiff method, and b) user ranking of the three techniques.

In addition the overall user's impression of Mid-Air Hand and Mid-Air Phone is that they are very attractive (Attractiveness ATT> 1). For Touchscreen the overall impression is moderately attractive (ATT = 0.5).

User preference

We asked participants to rank the three techniques in order of preference. 7 participants rated Mid-Air Hand as the preferred technique, 4 the Mid-Air Phone and 1 the Touchscreen pad. 7 participants ranked Touchscreen pad in last position. For the statistical analysis, we marked the most preferred technique with one and the least preferred technique with three. A Friedman test did not reveal any significant effect on the mark representing the user preference on interaction technique ($\chi^2(2)=5.17$, $p=0.075$).

Finally the most frequently mentioned positive comments refer to an accurate, funny and intuitive technique for Mid-Air Hand (P1, P3); an intuitive and easy to use technique for Mid-Air Phone (P3, P5); and a familiar technique for Touchscreen (P6, P12). The most frequently mentioned negative comments relate to the weight of the smartphone and the fatigue for Mid-Air Hand and Phone (P1); the lack of accuracy with distant targets and the loss of reference with quick movement for Mid-Air Phone (P3, P10); and the lack of accuracy, the two-handed aspect and the difficulty to combine movements in 3-DOF for Touchscreen (P6, P5).

Summary

Our study reveals that techniques based on direct mapping (Mid-Air Hand and Phone) are better than those based on indirect mapping (Touchscreen) for controlling the 3D translation of a Detail view. The study also reveals Mid-Air Hand scores better in terms of attractiveness and user preference, although there is no significant difference concerning SUS score. An interesting result of our study is that Touchscreen, i.e. the most common technique, is the worst in terms of performance, of perceived attractiveness and of user preference. These results are very encouraging and lead us to further explore the two mid-air techniques in a second experiment: the goal is to better investigate their use in the context of interaction with a public display.

EXPERIMENT 2: UNCOVERING THE GESTURES

The goal of this second experiment is to evaluate the difficulty of performing the two mid-air techniques in usual public context, i.e. without training and without human explanation. We thus study the impact of revealing strategies.

Task and apparatus

The task is the same as in the previous experiment: reach a target in a 3D space. We decided to display the 3D space without occlusion objects.

The apparatus is the same than in previous experiment: one sensor was attached to the smartphone while the other was attached to the hand.

Revealing strategies

Inspired by [43], we explore different revealing strategies. In the context of smartphone-based interaction with a public display, we focus on personal feedback on the smartphone. We study two different strategies to reveal the mid-air gestures: image (Figure 2) or a text explaining the gesture to perform. The text to explain the Mid-Air Hand was: "To move the view in the 3D scene, move your right hand in mid-air behind the smartphone while pressing the on-screen button". The text to explain the Mid-Air Phone was: "To move the view in the 3D scene, move the smartphone in mid-air while pressing the on-screen button".

Participants

We recruited 24 participants (4 females) from the local university, aged 27.54 years on average (SD=5.44). 8 of them had experience with 3D on mobile devices and 22 with mid-air gestures (Wiimote or Kinect). None of them participated in the previous experiment and we ensured they had not heard of it.

Design

To evaluate the different revealing strategies, our experiment followed a 2x2 between-participants design with Technique (Mid-Air Hand or Mid-Air Phone) and Revealing strategy (Text or Image) as factors. Every participant performed the two Techniques using one of the two Revealing strategies only. We counterbalanced the order of the Technique factor across participants: half of the participants started with one technique and half with the other. For every technique, each participant performed 8 selections on randomly selected targets (all at the same distance from the initial position).

Procedure

We decided to conduct a controlled experiment in order to exclude the imitation effect of public context [43], i.e. to reveal the gestures to future participants. Moreover the apparatus implied that participants could not perform the tasks with their own mobile device, excluding a large field study in a public space.

The study was performed in the presence of the interviewer. Participants were equipped with the sensors and explained the task (reach a target) without describing the interaction techniques. We informed them of the contextual revealing message (text or image) that would describe the technique. Participants watched the revealing message and then performed the task eight times for each technique. In case the participant took more than 2 minutes to understand the technique (i.e. perform the first trial), we ended the technique's block and marked it as a failure. This time (120 s) has been identified as the maximum time a user will try to perform an interaction in public spaces [43].

To diminish the influence of the sensor on the hand (that could partially reveal the nature of the Mid-Air Hand gesture), both sensors were attached during all the experiment. We did not inform users of what these sensors were used for.

Collected data

We logged all tracking data and measured time to complete the task from stimulus onset (including the time to read the help message). We measured user preference and perceived difficulty using a 5-points Likert scale.

EXPERIMENT 2 : RESULTS

Some participants failed to use some techniques since they were unable to understand the gesture. We first report on success rate and on data collected before analyzing quantitative and qualitative results.

Success rate and data collected

Among the 24 participants, 2 of them (8.3%) did not understand the Mid-Air Hand gesture and 1 (4.15%) the Mid-Air Phone gesture. These 3 participants were unable to understand in less than 2 minutes the technique, which was the first one to be used in the experiment. The Revealing strategy of these three failed conditions was 1 Text and 1 Image for the two Mid-Air Hand failures, and 1 Image for the Mid-Air Phone failure. Once these participants failed to perform the technique (always the first one), we asked them to perform the second technique while still providing the same revealing message: they all succeeded.

Consequently we collected a different number of trials for these three participants than for the others. For the 3 participants that failed to perform one technique we collected 8 trials (1 technique x 1 revealing strategy x 8 repetitions) x 3 users = 24 trials. For the 21 other participants, we collected 16 trials (2 techniques x 1 revealing strategy x 8 repetitions) x 21 users = 336 trials. In total we collected 360 trials.

Quantitative results

Task completion time

A Shapiro-Wilk test shows that the task completion time data do not follow a normal distribution ($p < 2.2 \times 10^{-16}$). None of the data transformations we tried allowed us to use parametric tests. The statistical analysis is thus based on non-parametric tests. We first compare for each technique the task completion time between trials when the technique was performed first or second. A Mann-Whitney test indicates that the order of techniques does not have a significant effect on task completion time (Z=6.62, p=0.2755 for Mid-Air Hand used first and Z=6.0, p=0.7956 for Mid-Air Phone used first). Thus we use all collected trials (including the 3 participants that failed one technique) for our analysis.

A Mann-Whitney test reveals the overall difference between the two techniques is not statistically significant (Z=9.0, p=0.38). The average task completion time for Mid-Air Hand is $24.54s$ ($SD = 34.85$) and for Mid-Air Phone $21.50s$ ($SD = 26.57$). We refine this analysis to distinguish the results obtained for each of the 8 trials performed by participants (Figure 7). For the first trial a Mann-Whitney test reveals a significant difference between the two techniques (Z=4.73, p=0.04), Mid-Air Hand being slower ($95.55s$ $SD = 53.71$) than Mid-Air Phone ($63.39s$ $SD = 43.99$). But for the other seven trials the difference is not significant (Figure 7). We note that the results obtained for these seven trials match the measures observed during the training session of experiment 1 (on average 81 second for five successfully repeated trials).

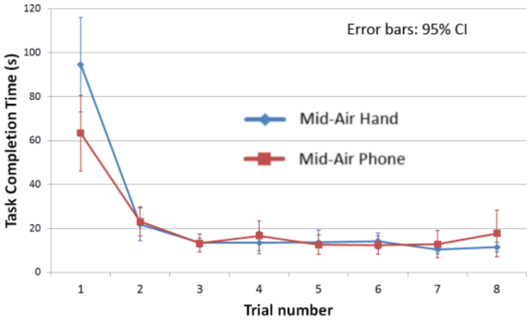


Figure 7: Task Completion Time (in s) for the 8 trials

Effect of the revealing strategies

A Mann-Whitney test reveals there is no significant effect of the revealing strategy factor (Z=4.71, p=0.35 for Mid-Air Hand and Z=5.08, p=0.50 for Mid-Air Phone).

Qualitative results

Both techniques get similar results in terms of perceived difficulty: 58% of participants find the Mid-Air Hand easy to use ("agree" or "strongly agree") vs. 63% for Mid-Air Phone. A Mann-Witney's test shows that there is no significant effect of the technique on the perceived difficulty expressed on the 5-Likert scale question (Z=-0.13, p=0.90). Interestingly, user preference produces different results for both techniques: 71% of participants like the Mid-Air Hand ("agree" or "strongly agree") whereas only 46% like the Mid-Air Phone technique (Figure 8). A Mann-Witney's test shows a significant effect of the technique on the overall user's rating expressed on the 5-Likert scale question (Z=2.23, p=0.026). This last result permits to reinforce the trend highlighted in the experiment 1 with regards to the hypothesis H3: user prefers interacting with Mid-Air Hand than with Mid-Air Phone.

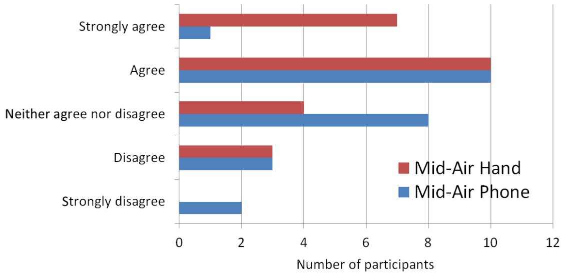


Figure 8: Likert scale results on the user preference for the two interaction techniques

Summary

Overall, results confirm that mid-air gestures can effectively be used to interact with O+D interface on 3D public display: 91.7% of participants have successfully used the Mid-Air Hand technique and 95.8% the Mid-Air Phone without any training or human explanation. These percentages would probably rise in a public context since users would be able to imitate other participants as observed in [43], thus enhancing the overall understanding of the gesture to perform.

A surprising outcome of our study is that the Mid-Air Hand gesture is more difficult to understand and to perform at first than the Mid-Air Phone. Not only the success rate is higher for the Mid-Air Phone, but it also allows a faster interaction during the first trial. However, our results also reveal that after the first trial, both techniques are comparable in terms of task completion time. Interestingly our study shows that despite the initial difficulty, participants preferred the Mid-Air Hand technique.

DISCUSSION

Based on the results of our two experiments we draw a set of design guidelines for smartphone-based interaction with Overview+Detail (O+D) interfaces on 3D public displays.

Gesture type and spatial mapping

One of the main findings of our experiments is that mid-air gestures are more efficient and preferred than touchscreen input for 3D interaction. Previous works on mid-air interaction with 2D content showed mixed results: some found mid-air interaction to perform as well as touchscreen [26], while others found touchscreen input to perform better [30]. Our results indicate that mid-air interaction is the best solution for interacting with 3D content. This result can be explained by the straightforward mapping between the gestures and the 3D translation.

Mid-air gestures reference

Our study establishes that with some training relative gestures (Mid-Air Hand) are as efficient and more preferred than absolute gestures (Mid-Air Phone). This result is in line with previous work on the use of spatial references for gestures [15,20]. One drawback is that the length of user's arms limits the Mid-Air Hand interaction. To overcome this limitation, the physical space behind the smartphone could be split into two areas: hand movements in the closest area would drive a position control interaction while movements in the furthest area would drive a rate-based control interaction, as explored in [41].

Gestures according to expertise

Our study also highlights that Mid-Air Hand performance depends on the user's expertise. To understand the gesture and perform it correctly, a novice user with only a text or an image as instructions is facing more difficulties than with the Mid-Air Phone (Figure 7). However, after the first trial users' performances are similar. The novelty of the Mid-Air Hand interaction, which is still far from being established, can explain this result.

Gesture information

A direct implication of the previous finding is that designers of 3D public applications should pay attention to the message used to reveal the interaction. Based on the experiments reported in [43] on 2D public displays and gestures, our second experiment compared the use of text and image revealing messages. As opposed to [43], we did not observe any significant differences. Given the unusual interaction, a more concrete instruction (e.g. video based or a combination of image and text) might be more appropriate. These results motivate us to keep investigating in the future the use of other revealing messages to explain how to physically operate a 3D mid-air input interaction.

Differences between our study and public context

Our second experiment was a controlled study representing a worst-case scenario. This experiment consisted in asking a user to perform a gesture never seen before. In a public place like a museum, the visitor is not alone and chances are s/he will observe others interacting with the system. The user will then benefit from an imitation effect [43] and the Mid-Air Hand performance will probably enhance.

To validate the interest of mid-air gestures for interacting with 3D O+D interfaces, we have employed the two mid-air techniques (Mid-Air Hand and Mid-Air Phone) in a concrete case study (Figure 1-b, c). The 3D scene, projected on a public display in the local university hall, represents a large telescope. The goal is to explore the different parts of the telescope and understand how it works. During two days, a large and varied audience (approx. 100 visitors composed of students, teachers and external public) virtually explored the dome of the telescope. This in-situ evaluation enriched the feedback from our experiments and permitted us to identify some limitations about the techniques.

Limitations

The addition of a sensor on the mobile phone and the hand makes them slightly heavier. As a result, several participants pointed out some muscle fatigue when using the Mid-Air Phone and Hand techniques. The size and weight of the mobile device should be limited to take fully advantage of these techniques. In the future, we will consider a better use of the embedded sensors to address this problem and the use of vision tracking to detect the hand position on the Mid-Air Hand technique.

To perform the Mid-Air Hand technique, we used the back area of the device as in [27]. This area was established after a preliminary study aimed at selecting the optimal width, height and depth of the mid-air interaction area. During the experiments, some participants wished to perform larger gestures. It should be possible for a user to calibrate his hand movement or to adjust the interaction area. In the future, we plan to consider other areas such as the side of the device as in [26].

Finally, the in-situ installation revealed a limitation when selecting an object in the 3D scene with the Mid-Air Hand technique. Newcomers were facing problems when trying to validate the selection with the hand handling the mobile phone instead of the hand behind the phone. This sometimes resulted in very difficult thumbs motions, especially when the target was close to the border of the screen. Using the hand behind the device is neither optimal, as it induces mid-air clutching. Designing different selection procedures, such as finger gestures (pinching for instance) or device actions (such as pressing a physical button), should be considered and offered as an alternative to touchscreen input.

CONCLUSION AND FUTURE WORK

In this paper, we explored the design space of mobile-based interaction with Overview+Detail (O+D) interfaces on 3D public displays. We evaluated three mobile-based techniques from the literature to navigate in a 3D scene: mid-air gesture around the mobile device (Mid-Air Hand), mid-air gesture with the mobile device (Mid-Air Phone) and touchscreen input. Our two controlled experiments show that Mid-Air Hand and Mid-Air Phone perform better than touchscreen input. However, the Mid-Air Phone gesture is easier to understand than the Mid-Air Hand gesture in usual public conditions, i.e. without training and with only a text message or an image as revealing information.

In the future we plan to increase the degrees of freedom of the navigation task by integrating the 3D rotation. For instance the Mid-Air Hand technique could be extended using gestures similar to [27]. To better implement mid-air solutions, we plan to remove the additional tracker: we will explore the combination of integrated sensors (gyroscope, accelerometer) and camera-based detection. Finally, we will extend our research to consider the selection and manipulation of the 3D objects in the 3D scene.

ACKNOWLEDGMENTS

We would like to thank the OMP-TBL members and all the participants who generously shared their time for the study. We are also grateful to the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper.

REFERENCES

1. 3D Ancient Wonders, archeological reconstruction online virtual museum. 2014. *http://www.3dancientwonders.com/*. Accessed: 2014-05.
2. Alessio, P., & Topol, A. 2011. A Public 3D Visualization Tool for the Musée des Arts et Métiers de Paris. Entertainment Computing – *ICEC 2011*, 6972, 136–142.
3. Bangor, A., Kortum, P.T., and Miller, J.T. 2008. An Empirical Evaluation of the System Usability Scale. In Proc. of *IJHCI, 24*, 6, 574–594.
4. Baudisch, P., Good, N., Bellotti, V. and Schraedley, P. 2002. Keeping things in context: a comparative evaluation of focus plus context screens, overviews, and zooming. In Proc. of *CHI '02*, ACM, 259-266.
5. Benzina, A., Dey, A., Toennis, M. and Klinker, G. 2012. Empirical evaluation of mapping functions for navigation in virtual reality using phones with integrated sensors. In Proc. of *APCHI '12*, ACM, 149–158.
6. Boring, S. et al. 2009. Scroll, tilt or move it: using mobile phones to continuously control pointers on large public displays. In Proc. of *OZCHI '09*, ACM, 161–168.
7. Boring, S, Baur, D., Butz, A., Gustafson, S. and Baudisch, P. 2010. Touch projector: mobile interaction through video. In Proc. of *CHI '10*, ACM, 2287-2296.
8. Bowman, D., Kruijff, E., LaViola, J. and Poupyrev, I. 2004. 3D User Interfaces: Theory and Practice. Addison Wesley Longman, Redwood City, CA, USA.
9. Burigat, S., Chittaro, L., and Vianello, A. 2012. Dynamic visualization of large numbers of off-screen objects on mobile devices. In Proc. of *MobileHCI '12*, ACM, 93-102.
10. Burigat, S. and Chittaro, L. 2011. On the effectiveness of Overview+Detail visualization on mobile devices. *Personal and Ubiquitous Computing*. 17, 371–385.
11. Casiez, G., Roussel, N., and Vogel, D. 2012. 1 € filter: a simple speed-based low-pass filter for noisy input in interactive systems. In Proc. of *CHI'12*, ACM, 2527-2530.
12. Cockburn, A., Karlson, A., and Bederson, B. 2008. A review of overview+detail, zooming, and focus+context interfaces. *ACM Computing Surveys 41*, 1, 1–31.
13. Decle, F. and Hachet, M. 2009. A Study of Direct Versus Planned 3D Camera Manipulation on Touch-Based Mobile Phones. In Proc. of *MobileHCI '09*, ACM, 32–35.
14. Fu, C, Goh, W. and Allen, J. 2010. Multi-touch techniques for exploring large-scale 3D astrophysical simulations. In Proc. of *CHI '10*, ACM, 2213–2222.
15. Gustafson, S., Bierwirth, D., and Baudisch, P. 2010. Imaginary interfaces: spatial interaction with empty hands and without visual feedback. In Proc. of *UIST '10*, ACM, 3-12.

16. Grubert, J. Grasset, R. and Reitmayr, G. 2012. Exploring the design of hybrid interfaces for augmented posters in public spaces. In Proc. of *NordiCHI '12*, ACM, 238-246.
17. Hachet, M. et al. 2008. 3D Elastic Control for Mobile Devices. *IEEE Computer Graphics and Applications*. 28, 4, 58–62.
18. Hachet, M. et al. 2009. Navidget for 3D interaction: Camera positioning and further uses. *International Journal of Human-Computer Studies*. 67, 3, 225–236.
19. Hassenzahl, M., Burmester, M., and Koller, F. AttrakDiff. 2014, http://attrakdiff.de/index–en.html.
20. Hinckley, K., Pausch, R., Goble, J., and Kassell, N. 1994. A survey of design issues in spatial input. In Proc. of *UIST '94*, ACM, 213-222.
21. Hürst, W. and Bilyalov, T. 2010. Dynamic versus static peephole navigation of VR panoramas on handheld devices. In Proc. of *MUM '10*, ACM, Article 25.
22. Hürst, W. and Helder, M. 2011. Mobile 3D graphics and virtual reality interaction. In Proc. of *ACE '11*, ACM, Article 28.
23. Irrlicht Engine - A free open source 3D engine. 2014. *http://irrlicht.sourceforge.net/*. Accessed: 2014-02.
24. Ivy software bus: *http://www.eei.cena.fr/products/ivy/*. Accessed: 2014-02.
25. Jacob, R. J. K., & Sibert, L. E. 1992. The perceptual structure of multidimensional input device selection. In Proc. of *CHI '92*, ACM. 211–218.
26. Jones, B., Sodhi, R., Forsyth, D., Bailey, B., and Maciocci, G. 2012. Around device interaction for multiscale navigation. In Proc. of *MobileHCI '12*, ACM, 83-92.
27. Kratz, S. et al. 2012. PalmSpace: continuous around-device gestures vs. multitouch for 3D rotation tasks on mobile devices. In Proc. of *AVI '12*, ACM, 181-188.
28. McCallum, D.C. and Irani, P. 2009. ARC-Pad: absolute+relative cursor positioning for large displays with a mobile touchscreen. In Proc. of *UIST '09,* ACM, 153-156.
29. Müller, J., Jentsch, M., Kray, C., and Krüger, A. 2008. Exploring factors that influence the combined use of mobile devices and public displays for pedestrian navigation. In Proc. of *NordiCHI '08*, ACM, 308-317.
30. Nancel, M., Wagner, J., Pietriga, E., Chapuis, O., & Mackay, W. 2011. Mid-air pan-and-zoom on wall-sized displays. In Proc. of *CHI '11*, ACM, 177-186.
31. Nekrasovski, D. et al. 2006. An evaluation of pan & zoom and rubber sheet navigation with and without an overview. In Proc. of *CHI '06*, ACM, 11-20.
32. Ouilhet, H. 2010. Google Sky Map: using your phone as an interface. In Proc. of *MobileHCI '10*, ACM, 419-422.
33. Polhemus Patriot Wireless - Polhemus. 2014. *http://www.polhemus.com/motion-tracking/all-trackers/patriot-wireless/*. Accessed: 2014-01.
34. Rønne Jakobsen, M. and Hornbæk, K. 2011. Sizing up visualizations. In Proc. of *CHI '11*, ACM, 1451-1460.
35. Roto, V. et al. 2006. Minimap: a web page visualization method for mobile phones. In Proc. of *CHI '06*, ACM, 35–44.
36. Serrano, M., Hildebrandt, D. and Irani, P. 2014. Identifying Suitable Projection Parameters and Display Configurations for Mobile True-3D Displays. In Proc. of *MobileHCI '14*, ACM, 9 pages.
37. Song, P., Goh, W.B., Fu, C.-W., Meng, Q., and Heng, P.-A. 2011. WYSIWYF: exploring and annotating volume data with a tangible handheld device. In Proc. of *CHI '11*, ACM, 1333–1342.
38. Tangible Media Group. *http://tangible.media.mit.edu/project/tether*. Accessed: 2014-01.
39. Telkenaroglu, C. and Capin, T. 2012. Dual-Finger 3D Interaction Techniques for mobile devices. *Personal and Ubiquitous Computing*. 17, 7, 1551–1572.
40. Thorsten B, et al. 2006. Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. In Proc. of *AVI'06*, ACM, 233-240.
41. Tsandilas, T., Dubois, E., and Raynal, M. 2013. Modeless Pointing with Low-Precision Wrist Movements. In Proc. of *Interact '13*, Springer, 494–511.
42. Vajk, T., Coulton, P., Bamford, W., & Edwards, R. 2008. Using a Mobile Phone as a "Wii-like" Controller for Playing Games on a Large Public Display. *International Journal of Computer Games Technology*, 2008, 1–6.
43. Walter, R., Bailly, G., and Müller, J. 2013. StrikeAPose: revealing mid-air gestures on public displays. In Proc. of *CHI '13*, ACM, 841-850.
44. Zhai, Y., Zhao, G., Alatalo, T., Heikkilä, J., Ojala, T., & Huang, X. 2013. Gesture interaction for wall-sized touchscreen display. In Proc. *UbiComp '13*, ACM, 175-178.