Interactive Public Displays

3D Freehand Gestural Navigation for Interactive Public Displays

Gang Ren, Chuan Li, Eamonn O’Neill, and Phil Willis • University of Bath

The increasing use of large public displays in indoor and outdoor environments brings new requirements and opportunities for rich, engaging user experiences. There are many ways to interact with displays, such as with a keyboard and mouse or a touch surface.1 (For a look at other research on large interactive displays, see the related sidebar.) However, these devices are installed together with or close to the displays, so they’re often unsuitable for public displays. Such situations include displays in enclosed spaces or out of users’ reach, or large public displays intended to be viewed from a distance (see Figure 1). Another way to interact with displays is through mobile devices. However, connecting the devices to the displays normally requires specific configurations, so this approach might not be convenient or appropriate for users in public settings.

Gestural input is becoming increasingly popular, including both gestural input using hands-on devices (for example, the Wii Remote) and freehand gestural input using camera-based motion tracking (for example, Microsoft Kinect). As gestural input increases beyond the now common homegaming setting, freehand gestural interaction, which doesn’t require hands-on input devices, will likely become more important for interactive public displays.

In addition, with improvements in the ability to visualize urban environments in 3D, some applications, such as virtual tours of a historic site or navigation around a planned plaza, can benefit from 3D user interfaces and a more immersive user experience. However, most established interaction techniques are in 2D, so they might be unsuitable for these scenarios.

Here, we focus on using freehand gestures to navigate 3D visualizations. We designed gestural techniques and evaluated them in a formal, quantitative lab experiment and a more informal qualitative field study. The results have led us to several design lessons for freehand gestural navigation on large public displays.

The Quantitative Experiment

This experiment used gestural navigation to search a visualization of a 3D urban environment. To gain further insight into gestural navigation’s characteristics, we also compared it to keyboard-and-mouse navigation.

Design

Given our aim to use low-cost technology for freehand gestural interaction without users holding or wearing a device or fiducial marker, tracking small finger or wrist movements (for example, using a Kinect camera) would have been difficult. So, we used the acceleration or movement of larger body parts such as the hand or shoulder to control navigation.

Controlling movement. Previous navigation techniques with tracking devices provide valuable references for freehand-navigation design. For example, researchers have used Wii Remotes and other handheld tracking devices (see the sidebar, “Related Work in Navigation Techniques”). However, with current low-cost tracking cameras such as the Kinect uses, we couldn’t accurately track the hand’s pitch
Interactive Public Displays

Kaikkuikaniemi and his colleagues discussed the increasing use of interactive digital displays and their effect on people’s urban lives.1 Mary Czerwinski and her colleagues presented an overview of large-display interaction,2 and Alex Olwal and Steven Feiner investigated an interaction design framework for large displays.3 Desney Tan and his colleagues demonstrated that users are more effective at 3D virtual navigation when using a large display.4 Large interactive displays can also help communication in information visualization.5 Myron Krueger and his colleagues investigated controlling computer graphics on large displays with users’ live video images in real time.6 They argued that using hands and fingers offers more than just a simple alternative to desktop pointing devices.

Researchers have used direct pen-based interaction with large displays in a brainstorming scenario7 and have used pen-based gestures to design 3D artifacts with large displays.8 Xiang Cao and Ravin Balakrishnan investigated a passive wand tracked in 3D for interaction with large displays, using different gestures and postures of the wand for input.9

References

Related Work in Large Interactive Displays

(a) (b) (c)

Figure 1. Public-display settings unsuitable for desktop interaction techniques: (a) a large display, (b) a display in a shop window, and (c) an out-of-reach display. Such situations offer opportunities for gestural interaction.

and roll. Consequently, navigation techniques designed for handheld tracking devices weren’t suitable for freehand gestural navigation. We used the same initial position for the entire navigation task, enlarging that position from a single point to a larger area to accommodate low-accuracy freehand tracking. We set a sphere-shaped initial position centered on the hand’s starting position. When the user’s hand reaches out of this sphere,
The visualization of urban environments has been an important application domain in computer graphics. One popular way to provide an immersive user experience is with 360-degree panoramas. For example, Google Street View enables users to virtually visit many areas by moving between “bubbles”—immersive 360-degree panoramas. However, the user experience of navigating a virtual city is usually less satisfactory in a desktop setting, owing to the small display range and unnatural interaction provided by 2D hands-on input devices such as a mouse and keyboard or a touchscreen.

To improve the navigation experience, researchers have employed physical interaction techniques, such as using eye, head, or body movements. For example, the SpaceActor project developed two short sticks (called NaviSticks) and one long stick (called the Kwisath) to navigate virtual environments. In particular, the Kwisath’s top end contained a target for tracking with six degrees of freedom (DOF). Users specified the Kwisath’s initial position by tapping its bottom end on the floor. Moving the Kwisath up and forward while pressing a button on it resulted in the corresponding navigation movement. Its movement range specified the speed; its rotation controlled the direction.

Another example employed a Wii Remote, with the joystick controlling speed, the pitch controlling elevation, and the roll controlling direction. Researchers have also used a 3-DOF flying broomstick; our field study employed the same metaphor (see the main article).

Physical navigation has also been shown to provide better performance in immersive information visualization for data analysis and target search. Beatriz Sousa Santos and her colleagues compared navigation with a VR system using a head-mounted display to navigation using a traditional desktop setup. Although users were generally satisfied with the VR system, most performed better with the desktop setup.

To improve navigation of a 3D city, Timo Ropinski and his colleagues proposed a constrained-navigation metaphor. Their approach determined the navigation path and altitude on the basis of the city road network. To address the problem of alternating between an overview, a medium-distance view, and a close-up view, they coupled navigation speed to the viewing height and angle to keep the visual flow constant.

However, most physical navigation techniques require users to have a device—for example, to hold a motion-sensing device (such as a Wii Remote) or wear fiducial markers for a visual tracking system. Although some initial research on freehand interaction exists, such as using freehand gestures for 3D menu selection, there still isn’t much published research on using freehand gestures for navigation.

**Changing the view direction.** For movement controlled by a single hand, the user’s position in the virtual 3D environment changes, but the direction in which the user is looking doesn’t. We could have used a single hand position to also control the view direction. That is, when the hand moved left or right, the movement and view directions...
Interactive Public Displays

would change synchronously. However, our pilot studies showed that this approach made navigation difficult.

Previous studies used the rotation of the hand or head to change the view direction. However, a single low-cost camera currently can’t reliably detect these movements. On the other hand, it can effectively detect the rotation of large joints such as the shoulders. Given that people turn their shoulders when they change their view direction in the real world, we also used shoulder turning to change the view direction. When the user’s shoulders turn left or right, the view direction turns left or right until the shoulders return to the straight-ahead position (see Figure 2b).

Room for improvement. Unfortunately, users can’t look up or down. This is mainly because using a single inexpensive camera to track body movement isn’t sufficiently accurate or fast to capture such movement. So, introducing pitch could make navigation difficult to control.

We could also improve navigation performance by introducing different navigation techniques and settings according to the environment’s context and scale. Because this study was an initial investigation of freehand gestural navigation, we focused mainly on the basic design.

The Evaluation
Locating places of interest is a common task in geographic applications or virtual tours in 3D visualizations. Generally, users first find the target’s rough location on a satellite map or in a bird’s-eye view. To see the target more clearly, they navigate closer to it, typically moving to a street view. So, in this study, we investigated how different target distances and directions affect gestural navigation.

The 3D environment. Using CityEngine, we built a city of approximately 4 km² in an 18th-century Parisian style. It contained 11,507 buildings and 1,619 streets; the buildings’ average height was 20 m, and the street width varied from 8 to 16 m.

The task. Each target was a cube 1 m wide, 1 m high, and 1 m deep randomly colored red, yellow, or green. We placed the targets at various distances and in various directions from the navigation starting point. Each target was 5 m above the ground and in the center of a street that was 100 m long and 10 m wide (see Figure 3a). The starting point had a bird’s-eye view of the city.

The participants had to find a target and report its color. To require them to navigate down from the bird’s-eye view to the street view, we placed a 100-m² occlusion over the target. So, they couldn’t see the target unless they were close enough to the correct street. To help them find the target’s rough location, we provided a text description of it (front, back, right, or left) at the beginning of each trial. We also placed a yellow sphere with a 10-m diameter over the location, at the same height as the starting point. This enabled them to concentrate on navigation rather than on searching for the target location.

Navigation techniques. For the conventional interaction technique, we used free-look navigation with a keyboard and mouse. To control movement, the participants pressed the W (forward), S (backward), A (left), and D (right) keys. Normal navigation was at 37.5 meters per second; pressing the Shift key increased the speed to 750 m/sec. Moving the mouse left, right, forward, or backward turned the view left, right, up, or down.

For freehand gestural interaction, we used the navigation technique we described earlier. As we mentioned, we calculated the vector from the hand’s current position to its original position (when the trial started) to calculate the navigation speed and direction:

\[
V = \begin{cases} 
0 & (d \leq 50) \\
37.5 + 0.5(d - 50) & (50 < d \leq 200) \\
112.5 + (d - 50) & (200 < d \leq 350) \\
37.5 \times 20 & (350 < d)
\end{cases}
\]

where \( V \) is the navigation speed (m/sec.) in the digital space and \( d \) is the vector length (mm) in real space.

The view direction changed continually as the participants’ shoulders turned. The turn speed was

Figure 2. Our gestural-navigation design. (a) The hand controls movement; the vector direction determines the navigation direction, and the vector length determines navigation speed. (b) When the user’s shoulders turn left or right, the viewing direction turns left or right.
two times the shoulder turning angle each second. The turn action triggered only if the left and right shoulders' depth difference exceeded 100 mm.

**Independent variables.** The independent variables were Technique (keyboard and mouse, or gesture), Target-Distance (short was 400 m; long was 800 m), and TargetDirection (forward, backward, right, and left).

The equipment and setting. We used a Sanyo PDG-DWL2500 short-throw 3D projector at 1,280 × 720 resolution with a 203 × 115-cm projection screen and Nvidia 3D Vision glasses (see Figure 3b). With keyboard-and-mouse navigation, the participants sat in a chair 250 cm from the screen and used a Logitech wireless keyboard and mouse on a square table. With gestural navigation, we used a Microsoft Kinect camera with a refresh rate of 30 fps and the OpenNI API on Windows 7. The Kinect camera was 50 cm in front of the projection screen at a height of 90 cm. Participants stood 250 cm from the screen.

The participants. The nine male and three female participants ranged from 23 to 30 years old, with a mean age of 26.17 (standard deviation = 2.41). They were experienced computer users with some experience of gestural interaction, such as using a Wii Remote or Microsoft Kinect for games.

The procedure. We held one session for each navigation technique; half of the participants used the keyboard and mouse first. Each session comprised a practice block and two test blocks; each block had 24 trials (three trials for each TargetDistance and TargetDirection combination). For keyboard-and-mouse navigation, each trial started with a countdown and ended when the participant performed the confirmation action—in this case, pressing the space key to report the target color. For gestural navigation, participants reported the color by raising their left hand (they all were right-handed).

The Results
We analyzed the navigation time (the time from the first navigation action to the confirmation action), navigation distance, and percentage of movement in different directions. The program automatically logged the data. Some participants accidentally triggered the confirmation before they saw the target color; we ignored 12 such trials (1.04 percent of the total data).

Navigation time. We used a repeated-measures ANOVA (analysis of variance) for Technique × TargetDirection × TargetDistance. We found significant effects for TargetDirection ($F_{1,10} = 15.00, p < .001$) and TargetDistance ($F_{1,11} = 24.85, p < .001$). (Regarding TargetDirection, the sphericity assumption wasn’t met, so we applied the Greenhouse-Geisser correction; the corrected degrees of freedom are shown.) We found no significant effect for Technique ($F_{1,11} = 1.06, p = .33$). We found no interaction effects. For TargetDirection, post hoc Bonferroni pairwise comparisons showed that forward was significantly faster than right, left, and backward ($p < .01$). For TargetDistance, short was significantly faster than long ($p < .001$). The mean navigation time was 9.12 sec. ($sd = 3.00$ sec.)
for gestures and 7.85 sec. (sd = 5.25 sec.) for the keyboard and mouse.

**Navigation distance.** Again, we used a repeated-measures ANOVA for Technique × TargetDirection × TargetDistance. We found significant effects for TargetDirection ($F_{3,33} = 5.85, p < .01$) and TargetDistance ($F_{1,11} = 524.93, p < .001$). We found no significant effect for Technique ($F_{1,11} = 3.68, p = .08$). We found no interaction effects. For TargetDirection, post hoc Bonferroni pairwise comparisons showed that forward was significantly shorter than right, left, and backward ($p < .01$). For TargetDistance, short was significantly shorter than long ($p < .001$). The mean navigation distance was 696.18 m (sd = 249.42 m) for gestures and 650.55 m (sd = 222.78 m) for the keyboard and mouse.

**Percentage of movement.** With both navigation techniques, participants could move in different directions without changing the view direction. We analyzed the percentage of the total time for moving left, right, forward, and backward (see Figure 4). Analysis using a two-tailed dependent t-test found more forward movement with the keyboard and mouse than with gestures ($t_{11} = 10.51, p < .001$). Correspondingly, there was less movement in the other directions with the keyboard and mouse than with gestures: backward ($t_{11} = -5.03, p < .001$), left ($t_{11} = -8.78, p < .001$), and right ($t_{11} = -4.08, p < .01$).

**User feedback.** Two participants reported feeling uncomfortable using the mouse and keyboard but feeling better using gestures. This might have been because when they used the mouse and keyboard, the visualization moved along with the mouse and keyboard actions, whereas when they used gestures, the visualization moved along with their body movement. The latter was more similar to real-world interaction. Participants also commented that gestural navigation was more immersive and fun. Some participants said they preferred gestural navigation's speed control because acceleration was more incremental, whereas keyboard-and-mouse navigation offered only two speeds.

**Discussion**

Although using gestures was consistently slower than using a mouse and keyboard, this difference wasn’t significant. Given that keyboard-and-mouse navigation is often impractical in public spaces, freehand gestures could be a promising method, offering walk-up-and-use experience without any handheld or desktop input device. For people who feel uncomfortable using a keyboard and mouse for 3D navigation, gestural interaction might be more comfortable.

Gestural navigation had some interesting differences from hands-on navigation. For example, as we mentioned earlier, with gestural navigation, participants used more left, right, and backward movements. In addition, the target direction affected the navigation time more. This was partly because using gestures to control direction wasn’t as easy and accurate as using the mouse. However, to compensate, the participants exploited the greater freedom of 3D movement that freehand gestures offered. For example, the participants used more movement in the left, right, and backward directions with gestural navigation to compensate for the less accurate view direction control. The findings reinforce our view that it isn’t straightforward for designers of freehand gestural interaction to adapt their experience of traditional hands-on input devices.

**The Qualitative Field Study**

Lab-based controlled experiments always impose severe constraints on how much fun and how natural any experience can be. To gain further insight into these aspects of our interaction techniques, we conducted a much more informal study in which primary-school children played a game on a large display in their classroom.

**The Gestural Interface**

On the basis of the first study’s results and feedback, we modified the gestural interface. We used...
two hands to control the movement and view direction because changing the view direction might be easier and more natural with the hands than the shoulders. To make the controls easier to understand, we used a flying broomstick as the metaphor, which we found was a familiar concept for 3D navigation from books and films our participants enjoyed.

To make the broomstick easier to control, we added a handlebar toward the front (see Figure 5a). At the start of the game, players put their hands parallel and close to their waist. After the broomstick started flying, they used the midpoint between the hands to control movement in the same way as in the previous study. Turning the view direction was similar to steering a bicycle: the more players twisted their hands and moved toward one side, the faster the view direction turned. The broomstick had the same DOF as the previous study’s gestural interface.

The Evaluation

One version of the game used free-look navigation as in the first study; the second version used free-hand gestures. For gestural navigation, we calculated the speed and direction as in the first study (with the midpoint between the two hands replacing the single hand location). The view direction turn angle per second was the sum of the angle of the two hands’ twist ($\beta$ in Figure 5b) and the deviation angle of the hands’ midpoint from the body center ($\alpha$ in Figure 5b).

The task. Players navigated to hit ogres and avoid penguins. A player needed to hit five ogres to successfully finish the task; if the player hit two penguins, the task was failed. A $180 \times 180 \times 360$-m space comprised 54 cubes, each $60 \times 60 \times 60$ m and containing one ogre or penguin at a random location. We randomly allocated the ogres and penguins; each cube had a one-fifth chance of containing an ogre and a four-fifths chance of containing an ogre. The navigation starting point was 180 m in front of this space.

The equipment and setting. The equipment and setting were almost the same as for the first study, except the equipment was in a primary school classroom rather than a laboratory. We replaced OpenNI with the Kinect for Windows software development kit so that the participants could take turns without the game needing a calibration pose. We turned off the stereoscopic 3D effect so that both the player and nonplayers could view the scene without 3D glasses.

The participants. The participants comprised 24 boys and 22 girls from 6 to 10 years old. They all had some experience of gestural interaction, such as using a Wii Remote or Microsoft Kinect for games.

The procedure. The participants gathered in the classroom about two minutes before their trials; they had been engaged in their normal activities before that. Their teacher randomly divided them into groups of four or five, depending on their schedules. The experimenter then demonstrated the navigation techniques and task.

Next, the groups took turns playing the two versions of the game. Each session lasted 10 minutes (determined primarily by the school’s schedule); the group members took turns playing until the time was up. We randomized the order of navigation technique for each group because we couldn’t control the number of participants in each group or the total number of participants.

The Results

Because we couldn’t control the number of participants in each group or the number of trials per participant with each technique, we report only descriptive data rather than analytical statistics. For gestures, the average navigation time was 24.61 sec. ($sd = 12.34$ sec.), and the success rate was 80.90 percent (72 out of 89 trials). For the keyboard and mouse, the average navigation time was 26.94 sec. ($sd = 12.33$ sec.), and the success rate was 93.16 percent (109 out of 117 trials).

We also asked the participants which technique they preferred. Twenty-four participants preferred gestures; 22 preferred the keyboard and mouse.

Some participants said that the gesture-based version let them navigate in a more natural way with body movement—for example, leaning forward to put their hand farther forward to move faster, and turning to control direction. They also commented that gestural navigation was “cool”
and that they had a lot of fun using it. However, some participants thought the mouse-and-keyboard version was easier to use because they were accustomed to using that interface to interact with standard computers.

Discussion
This study uncovered some shortcomings of our approach but also revealed a potentially powerful benefit of gestural input in social situations.

User performance. The participants finished more trials with the keyboard-and-mouse version. The main reason was that switching between players took longer with the gesture-based version. For example, before each trial, the player needed to make sure the Kinect camera was tracking him or her. Also, players often needed time to get used to using gestures to start the navigation.

Also, with the keyboard-and-mouse version, pressing physical buttons provided tactile feedback, so the players could control the navigation state (moving or stopping) more effectively. With freehand gestures, the players performed all control actions “in the air” without tactile feedback, and controlling the navigation state wasn’t so easy, leading to more errors.

A possible solution is to provide more feedback of the user’s body position and the broomstick’s navigation state. This would give users a better perception of their body movements and the broomstick’s state.

Group behavior and experience sharing. With gestural navigation, the participants interacted more with each other. The group members had more conversations, and the nonplaying members engaged more in the game and shared the experience with the current player. For example, when one player navigated toward a penguin, other members shouted to alert her and tried to guide her by speaking directions or even moving their own bodies as if they were controlling the navigation themselves. That player tried to perform according to the other group members’ contributions.

In contrast, with the keyboard-and-mouse interaction, the participants communicated less. Typically, the player simply finished the game, quietly concentrating on the display, keyboard, and mouse.

This is mainly because with gestural navigation, the nonplayers could observe and understand the player’s actions immediately from his or her hand or body movements, so they could share the player’s experience more directly. With the keyboard and mouse, the nonplayers had difficulty distinguishing between the player’s key presses, so less of the player’s experience was shared with them.

This observation suggests that freehand gestural interaction provides a more natural user experience, not only from the perspective of navigation control by a single user but also from the perspective of sharing the user experience within a social group in a public space. Our observation that the players adapted their gestural interaction according to the behavior of the other people in the same social space is consistent with Uta Hinrichs and Sheelagh Carpendale’s finding that the social context influenced gestures.

Design Lessons
In our two studies, keyboard-and-mouse and gestural navigation had similar performance and user preference. Considering that desktop input devices are often difficult to set up and use in public spaces and with large interactive displays, gestural navigation enabled by a single inexpensive camera is a promising interaction technique.

Furthermore, as we mentioned earlier, many of our participants commented that gestural navigation was fun and more natural. This suggests that gestural interaction can be suitable for displays used for exhibitions, planning, or commerce. For example, users could employ gestures to explore a virtual reconstruction of a historical site, evaluate plans for a building, or examine products in 3D.

For large public displays, gestural interaction can enhance the experience of not only the current user but also the people sharing that public space or activity. The user’s activity, behavior, and even emotions can be shared more effectively than with desktop interaction techniques. Bystanders can observe, understand, and appreciate the player’s experience better thanks to the natural interaction of freehand gestures. This suggests that freehand gestural interaction with public displays deserves further exploration.

Despite freehand gestural interaction’s potential, challenges exist for designers of public displays using it. For example, in both studies we noticed that some actions (for example, stopping and turning) were more difficult with gestural navigation. Possible solutions include providing more visual or auditory feedback or using additional input from other modalities—for example, speech. However, not all combinations will be effective in settings with considerable visual or auditory noise or where some modalities might be inappropriate.
Most research on gestural interaction has used fiducial markers or handheld devices for motion tracking. However, in many scenarios, especially interactive displays in public spaces, we can’t expect users to have a specific handheld device, a data glove, or fiducial markers. We aim to enable effective hands-free interaction without users holding anything, wearing anything, or attaching anything to their body. The user’s body alone can be an effective input device, enabling flexible interaction with public displays. We expect further research to show how far we can push such interaction’s use and effectiveness.

**References**


**Gang Ren** is a PhD student in the University of Bath’s Department of Computer Science. His research interests are freehand gestural interaction and 3D user interfaces. Ren received an MS in computer science and technology from Tsinghua University. Contact him at garry.ren@gmail.com.

**Chuan Li** is a research officer in the University of Bath’s Department of Computer Science. He explores computer vision and machine-learning algorithms for interesting graphics applications. Li received a PhD in computer science from the University of Bath. Contact him at cl.chuanli@gmail.com.

**Eamonn O’Neill** is a Reader in the University of Bath’s Department of Computer Science. His research interests are human-computer interaction, mobile and pervasive computing, social network analysis, and participatory design. O’Neill received a PhD in computer science from Queen Mary, University of London. He’s a Royal Society Industry Fellow. Contact him at eamonn@cs.bath.ac.uk.

**Phil Willis** is the director of the UK Centre for Digital Entertainment and Professor of Computing in the University of Bath’s Department of Computer Science. His research interests are color graphics, VR, computer animation, visual effects, and algorithms for manipulating film images. Contact him at p.j.willis@bath.ac.uk.

Selected CS articles and columns are also available for free at [http://ComputingNow.computer.org](http://ComputingNow.computer.org).