Comparison of Multimodal Interactions in Perspective-corrected Multi-display Environment

Ryo Fukazawa[1] Kazuki Takashima[1] Garth Shoemaker[2] Yoshifumi Kitamura[1] Yuichi Itoh[1] Fumio Kishino[1]

[1]Osaka University [2]University of British Columbia

ABSTRACT

This paper compares multi-modal interaction techniques in a perspective-corrected multi-display environment (MDE). The performance of multimodal interactions using gestures, eye gaze, and head direction are experimentally examined in an object manipulation task in MDEs and compared with a mouse operated perspective cursor. Experimental results showed that gesture-based multimodal interactions provide performance equivalent in task completion time to mouse-based perspective cursors. A technique utilizing user head direction received positive comments from subjects even though it was not as fast. Based on the experimental results and observations, we discuss the potential of multimodal interaction techniques in MDEs.

KEYWORDS: Multi-display environments, perspective-aware interfaces, gestural interaction, gaze, pointing

INDEX TERMS: H5.2 [Information interfaces and presentation]: User Interfaces - Graphical user interfaces

1 INTRODUCTION

A variety of new display configurations are currently being incorporated into offices and meeting rooms. Examples include projection screens, wall-sized PDPs or LCDs, digital tables, and desktop/notebook PCs. These multiple displays are often simultaneously used during work because we expect an efficient workspace with a large screen real estate. However, managing windows, reading text, and manipulating objects can become very complicated since multi-display environments (MDEs) include displays that can be at different locations from and different angles to the user. Therefore, much research has been devoted to establishing sophisticated and effective interfaces for MDEs. For example, Nacenta et al. showed that a perspective-aware interface for MDEs significantly and substantially improved user performance [17]. In their study, displays were stitched seamlessly and dynamically based on user viewpoints, and users interacted with the multiple displays as if they were in front of an ordinary desktop GUI environment operated by a mouse.

In addition, MDEs often provide three dimensional large flexible workspace; however, traditional 2D inputs by mouse or touch screens are insufficient for 3D workspace. Multimodal interaction techniques are promising in such workspaces. Gestures such as finger pointing may provide natural techniques to access distant multiple display surfaces. User eye gaze and head direction may also be effective approaches to interact with displays that are individually located at different positions and orientations. However, the effectiveness of these interaction techniques in MDEs has not been examined.

In this paper, we compare three multimodal interaction techniques and a perspective cursor technique in a perspective-corrected MDE. Experiments revealed the benefits of multimodal interaction techniques in a MDE. The following are the main contributions of this paper:

- Developing multimodal interaction techniques in a MDE
- Comparing multimodal interaction with traditional mouse-based interaction and discussing the possibilities of multimodal interactions

2 RELATED WORK

2.1 Systems Utilizing Multi-Display Environments

Earlier research into multiple display systems focused on personal displays for individual users and large displays for shared use with the goal of optimizing collaborative processes [14]. Recently researchers have investigated more flexible interfaces integrating combinations of displays, such as tabletops, personal displays, and large vertical displays [4, 16, 24, 25, 26].

These MDEs provide graphical interfaces based on standard 2D WIMP paradigms with extended functionality such as cross-display operation [4, 10, 24], which is a new technique to manipulate objects [26], and replication techniques to access the proxies of distant content from more accessible locations [13, 33]. The geometry used in these systems is mostly inherited from desktop monitor interfaces. That is, rectangular elements are shown that assume the user is viewing the display perpendicularly. In 3D MDEs, however, displays are positioned at different locations and at a variety of orientations; a user will frequently be viewing a display from an oblique angle. The violation of the perpendicular assumption complicates the task of viewing, reading, and manipulating information due to perspective distortion. This problem might become serious when a user performs a complicated task in MDEs, and several solutions are explored.

E-conic, one promising solution, uses perspective correction on distorted display elements such as windows or cursors shown in three-dimensionally constructed MDEs [17]. The system includes middleware that enables seamless combination of several displays and the construction of three-dimensional display spaces. The benefit of correcting perspective in MDEs was investigated in a controlled experiment that compared perspective windows to flat windows on five basic interaction tasks. Results show that when using perspective windows, performance improved between 8% and 60%, depending on the task. This suggests that obtaining 3D positional information and correcting for perspective offers clear user benefits in multi-display environments.

Yamadaoka 2-1, Suita, Osaka, Japan.
hi-ci@hi-mail.ise.eng.osaka-u.ac.jp.

IEEE Symposium on 3D User Interfaces 2010
20-21 March, Waltham, Massachusetts, USA
978-1-4244-6847-8/10/$26.00 ©2010 IEEE

2.2 Interaction Techniques for Distant Displays

As screen real estate increases through the adoption of multiple displays and large wall displays, users obtain larger workspaces. However, traditional 2D input by mouse or touch screen is not optimal in such workspaces. It is difficult to navigate over very large distances using either touch or a mouse.

Ray type pointing with fingers, laser pointers, or other input devices is one effective interaction method for large distant display spaces [5, 20, 21, 35]. These input techniques are natural because people often point at objects in the real world using a finger or laser pointer [5, 20] and can indicate objects located at distant places without having to move [21, 35]. Additionally, these techniques do not require any physical surface on which to operate (as opposed to a mouse) [32, 34].

In addition, various other interaction techniques for large displays have been proposed. For example, Pick and Drop enables direct interaction with physical objects [27], and the Radar technique uses a miniature that represents the environment [19]. Several techniques improve on existing work. Such techniques include a cursor that brings objects closer [3], a cursor that extends to a distant area [26], and a throwing metaphor operation to move objects to a distant location [9].

2.3 Multimodal Interfaces

People communicate using multiple modalities in daily life. In particular, we effectively use such non-verbal communication channels as body/hand gestures, eye contact, and so on. Therefore, many research efforts have focused on developing multimodal interfaces.

Multimodal inputs have been used for many years. Bolts et al. proposed a system that allows users to create and edit figures using both gestures and voice commands [5]. Similarly, Fukumoto et al. proposed the Finger-pointer that supports gestural commands by hand [8].

Many example exist using gaze input as another representative type of non-verbal interaction. MAGIC pointing [37] supports the selection of icons using gaze information to control a cursor. Gaze-Orchestrated Dynamic Windows [6] and EyeWindows [7] allow users to select a window by staring for a few seconds. Although errors in gaze input are significant, experimental results indicate that users can select objects more quickly with gaze input than by mouse [30].

Voice input is a crucial topic since voice commands are effective when combined with other modalities like gestures [5]. Schmandt proposed a system that performs complex operations by analyzing voice [29]. Individual properties of a voice, such as loudness and pitch, can also be used as a signal [11].

As described above, multimodal interactions offer potential as human-computer interfaces. However, little research has focused on multimodal interfaces in MDEs

In this paper, we propose multimodal interaction techniques such as gestural ray pointing by finger or laser pointer for manipulation in multi-display environments. Although previous work found that ray pointing techniques are inferior to a mouse or other devices [32], its performance may be superior in 3D MDEs. We conclude this based on the assumption that gestural techniques with ray metaphors are natural ways to interact with 3D environments, and support easy to access to distant multiple display surfaces. Similarly, user eye gaze and head direction may also be effective approaches to interact with displays that are individually located in different positions and orientation. Gaze can move easily from display to display and depends on the user's attention and indicates user's purpose, such as the object to be moved or the location of a window of interest. For these reasons, we investigated multimodal interactions with three-dimensionally constructed MDEs.

3 Utilizing Multimodality in Three-dimensional Multi-Display Environments

As reviewed above, MDEs hold promise for making large, 3D information spaces because each display can be located three-dimensionally, and the information space can be altered by replacing or repositioning the display. In such environments, perspective window [17] is a powerful tool for allowing users to see information without distortion. In this study, we used perspective window, perspective cursor, and E-conic middleware [28] as an MDE platform because they can seamlessly combine several displays located at different positions and angles. They also enable us to construct 3D working spaces. In addition, perspective corrections allow interaction with multiple displays as if users are in front of large, ordinary desktop GUI environments.

Thus, in this paper, we developed multimodal interactions for three-dimensional MDEs based on our relevant work called E-conic.

3.1 Perspective Correction in Multi-display Environments

E-conic seamlessly combines multiple displays and the perspective correction of windows and cursors to reduce the distortion of viewing or manipulating objects in displays located at various positions and angles. In the following sections, we briefly describe the perspective correction of windows and cursors in MDEs, which are the main benefits of E-conic.

3.1.1 Perspective Window

The perspective window provides an undistorted view of information in MDEs regardless of the angle of the display [17]. It displays the same kind of contents as traditional 2D windows (e.g., a web browser or a text processor) but offers extra features derived from the system's perspective-aware capabilities. Perspective windows are rendered using a virtual plane that is perpendicular to the user in the center of the window, and that is then projected geometrically onto the display. If a window is simultaneously displayed across more than one surface, the perspective reduces fractures. Perspective windows also reduce representational disparity by eliminating the perspective distortion that affects windows located on different displays, simplifying comparisons between content.

While the rendering of the windows is governed by the user's viewpoint, measured by a 3D position sensor, they remain attached to a pixel in a display through an anchor situated in the top left corner of these GUI objects. Their shapes and orientations vary if the display or the user moves, but the objects remain attached to the same physical point in the display.

3.1.2 Perspective Cursor

The perspective cursor [18] is used by the system for cross-display manipulation, and works as follows. The 3D position coordinates of the user's head are measured, and a 3D model of the whole environment is simultaneously maintained with the physical position of all screens. The point-of-view coordinates of the user's head, obtained from the model, lets us determine which displays are contiguous in the field of view, which can be different from which displays that are actually contiguous in 3D space. The pointer's position and movement are calculated from the viewpoint of the user, who perceives the pointer movement across displays as continuous, even when the pointer's actual movement (considered in three-dimensional space) is not. The pointer travels through empty space to get from one display to the next. The cursor can in fact be in any position around the user, even if there are no displays to show its graphical representation. Few environments exist in which users are completely surrounded by displays, meaning that users might lose the pointer in non-displayable space. The solution we implemented is a perspective variant of halos [2]. The system provides a perspective halo based on the user's viewpoint. Halos are circles centered on the cursor whose radii appear, at least partially, on at least one of the screens. By looking at the circle's displayed parts (its position and curvature), the users can tell how far and in which direction the perspective cursor is travelling. When the cursor is barely out of one display, the halo's displayed arc section is highly curved, showing most of the circle. If the cursor is very far away, the arc will approach a straight line.

3.2 Developing Multimodal Interactions

In MDEs, user workspaces are much larger than traditional ones. Additionally, the location, orientation, and size of each display may be configured differently depending on the purpose of a user's work. This means that user can work more efficiently and variedly, but at the same time the user's tasks become are often more difficult and more time-consuming. For example, a user may need to move a mouse a long distance or across displays to click on a desired target object located far from the current cursor position. Users may lose the cursor when it enters a gap between displays. Accordingly, traditional two-dimensional inputs by mouse or even the perspective cursor have difficulty in MDE interactions that are constructed in three-dimensional spaces. MDEs need a better input technique that enables natural, explicit, and three-dimensional object manipulations. For these reasons, we believe that interactions exploiting user multimodality are applicable to multi-display environments.

Although many interactions employing users' multimodalities can be considered, we designed the following three interactions for MDEs with perspective windows.

Gestural interaction based on laser pointer. We developed a gestural interaction based on the laser pointer technique [20]. This technique utilizes the behavioral channel of the user's modalities and a metaphor of ray casting. Fig. 1 shows an example in which a user moves a cursor to the position indicated by gestural pointing. Gestural interaction with a ray, which seems to be a natural way for indicating a point on a distant display surface in three-dimensional spaces, is likely easy to understand, and can be used for manipulating cursors or objects. Fig. 3(a) shows a device for detecting the user's hand position and direction that consists of a stick with two buttons and a 3D positional tracker. By pushing the buttons, users can select an object and move (drag) it to another location.

Gestural interaction based on image-plane. We developed another type of gestural interaction based on image-plane [23]. This technique utilizes both user behavior and point-of-view. The cursor position on display is determined by the angles based on user point-of-view and hand (Fig. 2). With this technique, users can easily determine the cursor's position, because its position on the distant display almost corresponds to the position of the user's moving hand from the user's view. Our implementation of this technique uses the same device to detect user's hand position and movement. The user's point-of-view is determined from head position and orientation, measured by a 3D tracker attached to a hat (Fig. 2).

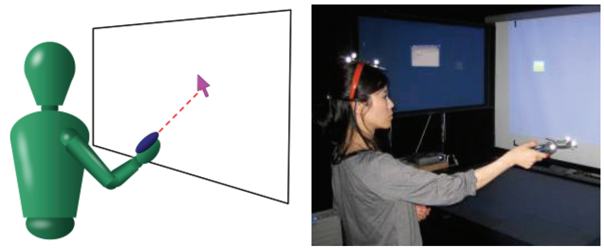

Fig. 1: Gestural interaction based on laser pointer

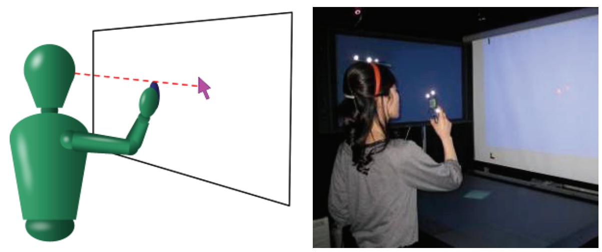

Fig. 2: Gestural interaction based on image-plane

(a) Gesture (b) Head

Fig. 3: Devices for detecting gesture, head position and direction, and eye gaze

Interaction based on head/eye direction (gazed point). We also developed an interaction technique utilizing the head/eye direction detected by a 3D tracker. The technique supports cursor jumping to the position of the user's head/eye direction point (this technique is based on MAGIC pointing [37]) and user also can move cursor normally by using mouse. Cursor jumping with head/eye direction prevents the user from losing the cursor. Fig. 3(b) shows a device for detecting the head direction of a user that consists of 3D positional trackers.

These three interaction techniques were designed to ease manipulation of a cursor, especially for longer target distances. The head/eye inputs can also be effectively used for several other purposes, particularly head/eye movements that can be registered as often-used actions. Although there were other possibilities using voices, we focused on three interactions based on head tracking and gesture input in this study because voice can be used for trigger commands, but it needs to be used with a direct input method like gestures or mouse for such object manipulation tasks as dragging. We used head direction as a proxy for the user's gaze because commercial eye-tracking systems do not support moving users, and users of MDEs are often in motion. We consider this compromise acceptable, as other systems have used head direction in a similar fashion [1].

4 Empirical Study: Comparison of Multimodal Interactions in Multi-display Environment

In Section 3, we proposed three multimodal interfaces that can be adapted to three-dimensionally assembled multi-display environments.

We experimentally compared the proposed three multimodal interaction techniques and the perspective cursor. We chose to evaluate perspective cursor instead of a traditional cursor because perspective cursor has been shown to perform better for pointing, steering, and reading tasks. The following sections describe the apparatus, task, conditions, design, measurement method, analysis methodology, and experiment results.

4.1 Apparatus

An MDE was set up with four displays: a bottom-projected tabletop display (horizontal, 1000 x 750 mm, 1024 x 768 pixels), a large projected display (vertical, 1300 x 950 mm, 1024 x 768 pixels), another projected display (vertical, 800 x 600 mm, 1024 x 768 pixels), and a large PDP monitor (vertical, 1100 x 620 mm, 1024 x 768 pixels). These displays were set up in three-dimensional positions and orientations in front of a user who stood obliquely with an either an optical mouse or a pointing device, used for input to all displays (Fig. 4). The mouse was used on an extra large 408 x 306 mm mouse pad (Power Support's Airpad Pro III, AK-07).

Head position/orientation and hand position/orientation were tracked by an OptiTrack V100. The sensor was placed on the user's head and had to be worn throughout the whole experiment. Users stood in front of the displays and could move freely.

The entire setup was run on three machines: a PC (Intel(R) Core(TM) 2 Duo CPU E6850 3.00 GHz, 2.00 GB RAM) controlling the tabletop display, projector, and a modified E-conic server; a second PC (Intel(R) Core(TM) 2 Duo CPU E6850 3.00 GHz, 2.00 GB RAM) controlled the vertical projected display and the PDP monitor; and a third PC (Intel(R) XEON(TM) CPU 2.20 GHz, 2.00 GB RAM) ran the experimental software that rendered the contents of windows. The machines were connected by a Gigabit Ethernet network.

4.2 Task

Participants were asked to drag an object to a target window as fast and accurately as possible. The object and window were represented as a folder icon and a target window similar to those of a typical operation system. This dragging task was modeled after common operations in GUI systems. The display boundaries potentially cause a problem on the cursor or the object transitions across displays that are very important and fundamental operations in MDEs. Unlike the fundamental pointing task, since this task can simulate that users absolutely must move the object across displays (beyond their gap), evaluating the effectiveness of the multimodal interactions against the problem of the display boundaries is reasonable.

In all trials, the folder and the target window were located on different displays (tabletop-projector, projector-PDP, projector-projector etc.), but they were always oriented towards the user using the perspective window. Users moved the folder from one display to another. If the folder was successfully dropped in the target window, the folder and target window were re-located and the next trial began. If the user missed the target window then the trial was repeated until it was performed successfully.

4.3 Conditions

Techniques: We tested five primary techniques: 1) Perspective Cursor (normal CD gain): PC (normal); 2) Perspective Cursor (high CD gain): PC (high); 3) Gesture (laser pointer); 4) Gesture (image-plane); and 5) Head Direction based on MAGIC pointing. The details of these techniques were described in Section 3. We included two levels of cursor CD gain in order to observe the effect of CD gain in large three-dimensional display spaces. We defined the CD gain of the perspective cursor in terms of the distance the mouse was moved relative to the angle the cursor moved. This means that the cursor's visual speed from the user viewpoint remained constant in all displays, given constant mouse movement. In this experiment, we used three degree/mm as a normal CD gain and six degree/mm (double of normal) as a high CD gain. Fig. 5 shows actual experimental setting.

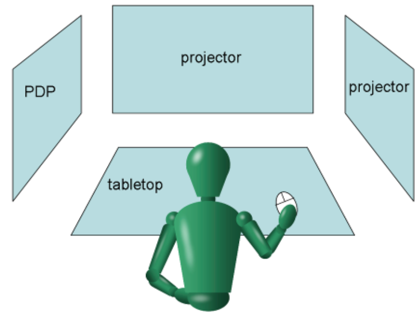

Fig. 4: Experimental setting

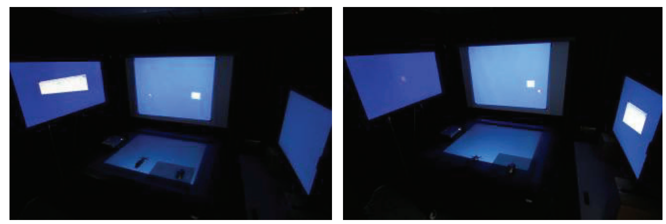
Fig. 5: Actual experimental settings

Distance between displays: All experimental tasks were performed on the four displays and every task required moving a folder between displays. We manipulated the distance between displays (d) in order to examine the impact of distance on user performance. In the first arrangement, the distances between displays were set to about 450 mm, and in the second, distances were set to about 900 mm. Fig. 6(a) shows the distances between displays.

Direction of tasks: All experimental tasks moved a folder across displays in four directions: horizontal, vertical, horizontal (long distance), and diagonal (Fig. 6(b)). The horizontal (long distance) task moved over two gaps of displays (90 degree when the d is 450 mm and 120 degree when the d is 900 mm), while the horizontal task only moved over one gap (45 and 60 degree).

4.4 Design

In the first display configuration (d = 450 mm), 12 volunteers (11 males, 1 female) from a local university, aged 21 to 25 (average 22.83) participated. In the second configuration (d = 900 mm), 12 volunteers (11 males, 1 female) from a local university, aged 21 to 25 (average 22.75) participated. All participants were right handed and manipulated the mouse or gestural input device with their right hand. All participants were tested individually and performed each task 64 times with each of the techniques. The order in which they performed the different conditions and the order in which the displays were used as targets and origins were balanced across subjects.

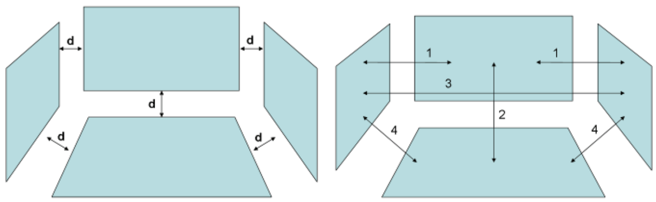

Fig. 6: Distance between displays (*d*), and directions of tasks (1-4): 1 is horizontal, 2 is vertical, 3 is horizontal (long distance), and 4 is diagonal.

4.5 Measures and Analysis Methodology

For the evaluation of each trial, we measured the time from the start of the drag to the final drop of the folder in the target window, and counted the drops on empty space as errors. Then we analyzed the throughput with methodology based on ISO9241-9 and Fitts' law research for pointing device evaluation [31]. Although Fitts' law was originally designed for 1D or 2D tapping tasks, it has been shown to be robust for non-planar pointing as measured in polar coordinates [12, 15, 36]. When applied in polar coordinates the only necessary change is to substitute angular measurements (degrees) for linear measurements (mm) in the standard Fitts' equation.
The following equations outline the throughput calculation approach.

$$ID = \log_2\left(\frac{D}{W} + 1\right)$$

$$TP = \frac{1}{y}\sum_{i=1}^{y}\left(\frac{1}{x}\sum_{j=1}^{x}\frac{ID_{ij}}{MT_{ij}}\right)$$

The first equation calculates the index of difficulty (*ID*), and the second is for the throughput (*TP*). *D* is the distance from the start position of the folder icon to the position of the target window (degrees), *W* is the width of the target window (degrees), *MT* is the task completion time (ms), and x and y are the number of movement conditions and participants. Although the original throughput formulation takes into account errors, we used a simpler formulation that ignores errors. This could be considered to be a "raw" throughput. We feel this is valid as few errors occurred in the dragging task.

4.6 Results

4.6.1 Throughput

A three-way ANOVA by technique, distance, and direction was carried out on the throughput. Fig. 7 shows throughput relative to the technique. We found a main effect of technique ($F(4,44) = 64.8858$, $p < 0.01$), distance ($F(1,11) = 72.0218$, $p < 0.01$), and direction ($F(3,33) = 53.7337$, $p < 0.01$). We found interactions between technique × distance ($F(4,44) = 4.8035$, $p < 0.01$) and technique × direction ($F(12,132) = 3.6417$, $p < 0.01$). A multiple comparison using Tukey's HSD test on technique shows that PC (high), Gesture (laser pointer), and Gesture (image-plane) were not significantly different from one another, and were significantly faster than PC (normal) and Head Direction. A Tukey test on the first interaction showed that the head direction of the 900 mm distance was significantly faster than the 450 mm distance. This means that the effect of the head direction increased with greater distance. Tukey's test on the second interaction revealed that both perspective cursors of the horizontal tasks were

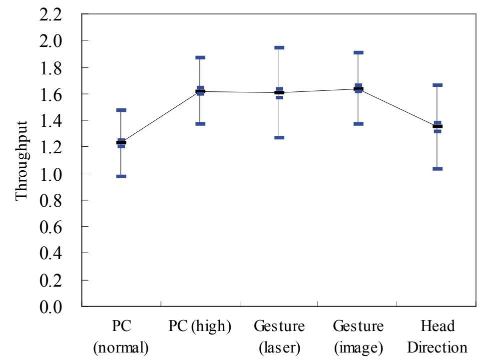

Fig. 7: Throughput relative to techniques

Table 1 Average subjective evaluation scores

	PC (normal)	PC (high)	Gesture (laser)	Gesture (image)	Head Direction
Preference	2.500	3.292	**3.792**	**3.792**	3.708
Location awareness	3.250	2.083	**4.042**	**4.250**	3.500
Physical fatigue	2.375	**3.208**	2.750	2.417	**4.292**
Intuitiveness	3.000	3.083	**4.292**	**3.958**	3.000

faster than those in other directions. In other techniques, there were no differences among the directions.

4.6.2 Number of Errors

A three-way ANOVA by technique, distance, and direction was carried out on the number of errors. We counted the dropped folder icons in empty space as errors. We found a main effect of technique ($F(4,44) = 14.4181$, $p < 0.01$). No main effect of distance ($F(1,11) = 0.0138$, $p = 0.9065$) or direction ($F(3,33) = 0.1105$, $p = 0.9540$) was revealed.
The worst average number of errors was 1.021 (number per trial) for Gesture (laser pointer), and this value was significantly different from the other four techniques by Tukey's test ($p < 0.01$). There were no differences among the other four techniques. The errors of PC (normal), PC (high), Gesture (image-plane), and Head Direction were 0.281, 0.302, 0.469, and 0.594, respectively. These observations suggest that Gesture (laser pointer) underperformed the others because its manipulation was strongly affected by the jitter of the 3D tracker and the shaking of the user's hand.

4.6.3 Subjective Evaluation

After completing all trials, the participants answered questionnaires. Four questions were answered on a one-to-five scale, related to preference, location awareness, physical fatigue, and manipulation intuitiveness. The average scores of all questions by technique are shown in Table 1.
An ANOVA by technique and distance was carried out on the four items. We found a main effect of technique in all items, preference ($F(4,44) = 4.8594$, $p < 0.01$), location awareness ($F(4,44) = 20.1457$, $p < 0.01$), physical fatigue ($F(4,44) = 11.1140$, $p < 0.01$), and intuitiveness of each technique ($F(4,44) = 9.0292$, $p < 0.01$). No main effect of distance was found in any items.

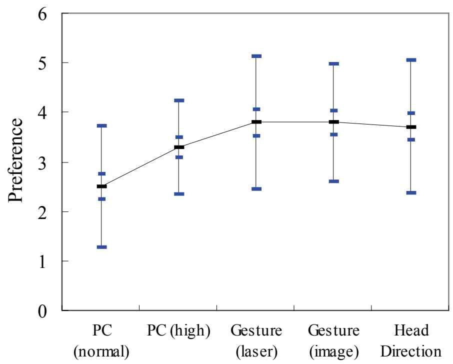

Fig. 8: Average preference scores relative to techniques. (5=best, 1=worst)

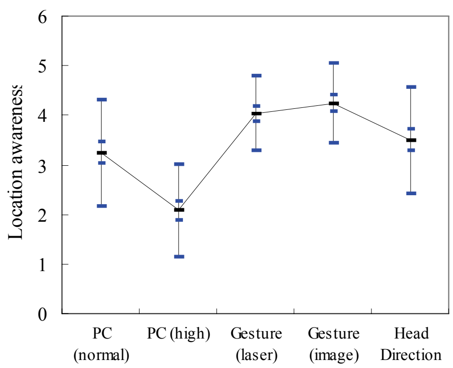

Fig. 9: Average location awareness scores relative to techniques. (5=best, 1=worst)

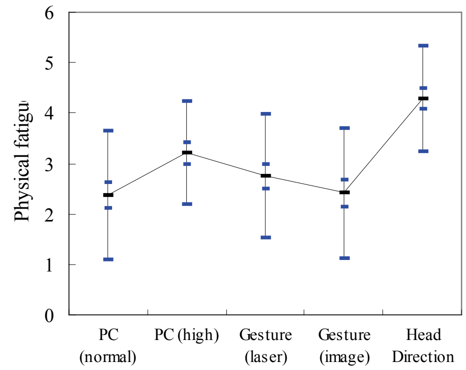

Fig. 10: Average physical fatigue scores relative to techniques. (5=best, 1=worst)

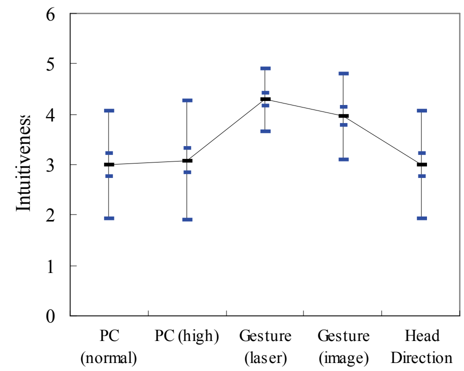

Fig. 11: Average of manipulation intuitiveness scores relative to techniques. (5=best, 1=worst)

Preference: Figure 8 shows the preference scores by technique. Tukey's HSD test revealed that PC (normal) significantly outperformed the others.

Location awareness: Figure 9 shows the location awareness scores by technique. Tukey's HSD test revealed that PC (high) was the worst ($p < 0.01$). PC (normal) was also outperformed Gestures and approximately equaled Head Direction.

Physical fatigue: Figure 10 shows the physical fatigue scores by technique. Tukey's HSD test revealed that the Head Direction was the best ($p < 0.01$). There was no significant difference among the other four techniques.

Intuitiveness of manipulation: Figure 11 shows the manipulation intuitiveness scores by technique. Tukey's HSD test revealed that both Gesture conditions significantly outperformed the others ($p < 0.01$).

5 Discussion

5.1 Comparison of Results

Throughput: Although previous studies with single display environments argued that gestural interactions are inferior to mouse-base interaction, our results with 3D MDEs showed that Gesture (laser pointer) and Gesture (image-plane) have approximately the same performances as PC (high) and outperformed PC (normal). Such enhanced performance of gestural interactions clearly increases their possibilities for object manipulations in 3D MDEs. Although, in this experiment, we counter-balanced to remove effect of familiarization and did not find a significant effect, we saw some participants perform better, especially with gestural interactions, after they had become familiarized. We hypothesize that any learning might be related to understanding the properties of the tracker. If participants understand the properties of the tracker, such as jitter or latency, they will perform better and be more accurate.

Furthermore, in this experiment, we used a camera-based tracker for gestural interactions that has larger latency than the optical mouse used in the mouse conditions. The theoretical value of the tracker's latency is about 10 ms, but it is likely larger in practice. According to Pavlovych's paper, tracker latency had a significant negative effect on laser-pointer based pointing throughputs [22]. Therefore our results might be affected by different device latency between gesture and mouse. However, the impact of additional latency in our gesture techniques can only result in performance being underestimated relative to the mouse conditions. It is therefore reasonable to reach our conclusion that the gesture techniques are as good as the mouse conditions. Furthermore, in the future tracker technology will improve and supported gestural interactions will become more usable and perform better.

Additionally, users frequently lost track of the cursor position in the perspective cursor with high CD gain condition. We hypothesize this to be due to poor visual feedback and excessively fast cursors. This drawback is universal to mouse techniques configured for large surfaces. On the other hand, gesture-based techniques support natural and effective interaction with distant and multiple displays without a visual feedback problem.

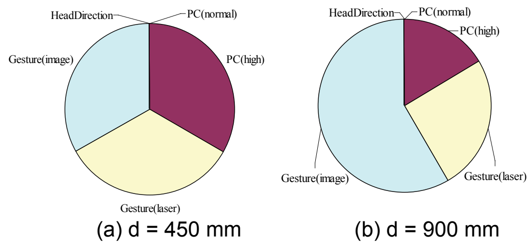

Fig. 12: Percentage of fastest technique

With perspective correction, the perspective window and perspective cursor provided a control space as if a user is in front of a traditional desktop GUI. Interestingly, contrary to past work on single display mouse input, horizontal tasks were found to have higher throughput than vertical tasks. We hypothesize this to be because a horizontal transition across display boundaries is a common operation in daily mouse use with dual-displays.

We did further throughput analysis by participant. Fig. 12 shows the breakdown of the fastest technique over all participants for both the short distance of 450 mm (Fig. 12(a)) and the long distance of 900 mm (Fig. 12(b)). By this measure the gestural techniques (yellow and blue areas) performed better in the environments with larger gaps. From this we conclude that gestural interactions are relatively more useful in MDEs with large gaps between displays.

Error: In this experiment, only Gesture (laser pointer) had a large amount of errors. User comments complained about tracker jitter and the shaking of users' hands. This is a potential problem when using 3D motion sensors. To solve it, we plan to revise the position detection algorithm to use a prediction filter or some smoothing filter.

Preference: There were no significant differences in preference among the highest scored three techniques. Since the perspective cursor has previously been shown to be a better interaction technique than a traditional mouse in MDEs, we expected that the perspective cursor would be preferred. Contrary to this hypothesis, the data shows that gestural techniques have comparable performance to the perspective cursor. This result highlights the effectiveness and the potential of gesture-based interactions in MDEs.

Location awareness: Although high CD gain results in high task performance, users frequently lost their cursors, especially when the cursor entered gaps between displays. Users also often overshot the target. The poor location awareness properties of the high CD gain cursor suggest a lower CD gain device in actual working situations may provide a balance between awareness and an ability to move long distances. On the other hand, gestural interactions have an advantage in location awareness, and users can easily find their cursor because of its explicitness. The best scored technique in awareness was gesture interactions based on an image-plane. The reason might be the high consistency between the cursor on the display and the movement of the user's hand from the user's perspective.

Physical fatigue: The technique with head direction causes less fatigue than the other four techniques because cursor jumping reduces the physical movement of the user's hand. On the other hand, gestural interactions tend to increase fatigue because they require movement of the user's whole arm and hand position or direction. This may be a problem when users need to do such operations for a long time and should be addressed in future work.

Intuitiveness of manipulation: Overall, gestural interactions achieved intuitiveness because they are based on user's natural behavior in daily life. In the future, a greater variety of displays and three-dimensional arrangement of workspaces might exist. In such environments, the intuitiveness of gestural interaction can be extended to more useful interactions.

5.2 Characteristics of Each Technique

Perspective cursor (interaction based on mouse): The higher the CD gain becomes, the more the task performance is improved. At the same time, however, users have a tendency to lose the cursor. This loss of cursor may cause users to overrun the destination, leading to wasteful movements and decreased location awareness. While performing natural interactions with a mouse can be difficult, this method has also an advantage of accurate and fast manipulation because it is not affected by the jitter of tracking hardware or the shaking of users' hands.

Gesture interaction: As a method, Gesture equals a mouse. Since it gives users intuitive operation, this scheme can be treated as one effective or beneficial interaction method. However, the precision of existing trackers is hampered by their mechanical limitations or by the shaking of the user's physical hands. Thus this class of techniques is applicable to object manipulation among displays in larger 3D MDEs but may not adapt to tasks that require a high degree of accuracy as painting or designing, at least until tracking technology improves.

Interaction with head direction: The task performance of this method is not as good as the other methods. However, since it produces limited fatigue effects it may work well in conjunction with other methods during long distance movements, or when a cursor gets lost. However, we need to carefully consider many identical comments by participants that had little burden in the task. In fact, this method is closer to head direction than eye gaze, so users needed to turn their necks too much. In the object placement task, users originally looked at the target. Using this action as the input, they might be relieved from controlling burdensome input hardware. Although some studies related to head/gaze-based-interface failed to show clear effectiveness, head/eye direction has potential in three-dimensional MDEs. However, one possible limitation of this system is the head movement direction. Thus these interaction techniques can be effectively used in middle-range MDEs that comprise of a few displays in about 100 degree of viewing angles.

6 Conclusion

In this paper, we developed multimodal interaction techniques for manipulation in multi-display environments and experimentally evaluated their performance. We found that gesture-based multimodal interactions provide approximately the same performance as mouse-based perspective cursors in task completion time. Overall gesture-based interaction received more positive comments than the mouse in location awareness and intuitiveness. The technique utilizing user's head direction offered high scores for reducing physical fatigue even though it was not fast. These results suggest that multimodal interactions might be effective in three-dimensional multi-display environments. In future work, we will explore interactions that include more modalities and adapt to larger 3D MDEs.

References

[1] M. Ashdown, K. Oka, and Y. Sato. Combining head tracking and mouse input for a GUI on multiple monitors. *Extended Abstracts on Human Factors in Computing Systems (CHI)*, pages 1188–1191, 2005.

[2] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing off-screen object. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 481–488, 2003.

[3] P. Baudisch, E. Cutrell, D. Robbins, M. Czerwinski, P. Tandler, B. Bederson, and A. Zierlinger. Drag-and-Pop and Drag-and-Pick: techniques for accessing remote screen content on touch- and pen-operated systems, *Proceedings of Interact 2003*, pages 57–64, 2003.

[4] J. T. Biehl and B. P. Bailey. ARIS: an interface for application relocation in an interactive space. *Proceedings of Graphics Interface (GI)*, pages 107–116, 2004.

[5] R. A. Bolt. Put-that-there: voice and gesture at the graphics interface. *Proceedings of Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 262–270 (1980)

[6] R. A. Bolt. Gaze-orchestrated dynamic windows. *Proceedings of Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pages 109–119, 1981.

[7] D. Fono and R. Vertegaal. EyeWindows: evaluation of eye-controlled zooming windows for focus selection. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 151–160, 2005.

[8] M. Fukumoto, K. Mase, and Y. Suenaga. Finger-pointer: a glove free interface. *Posters and Short Talks of Conference on Human Factors in Computing Systems (CHI)*, page 62, 1992.

[9] M. Hascoët. Throwing models for large displays. *Proceedings of HCI'2003*, British HCI Group, pages 73–77, 2003.

[10] K. Hinckley, G. Ramos, F. Guimbretiere, P. Baudisch, and M. Smith. Stitching: pen gestures that span multiple displays. *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI)*, pages 23–31, 2004.

[11] T. Igarashi and J. F. Hughes. Voice as sound: using non-verbal voice input for interactive control. *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 155–156, 2001.

[12] R. Jota, M. A. Nacenta, J. A. Jorge, S. Carpendale, and S. Greenberg. A comparison of ray pointing techniques for very large displays. *Report 2009-942-21, Department of Computer Science, University of Calgary*, 2009

[13] G. Khan, D. Fitzmaurice, N. Almeida, N. Burtnyk, and G. Kurtenbach. A remote control interface for large displays. *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 127–136, 2004.

[14] K. L. Kraemer and J. L. King. Computer-based systems for cooperative work and group decision making. *ACM Computing Surveys 20*, pages 115–146, 1988.

[15] B. A. Myers, R. Bhatnager, J. Nichols, C. H. Peck, D. Kong, R. Miller, and A. C. Long. Interacting at a distance: measuring the performance of laser pointers and other devices. *Proceedings of Conference on Human Factors in Computing Systems (GCHI)*, pages 33–40, 2002.

[16] B. A. Myers, H. Stiel, and R. Gargiulo. Collaboration using multiple PDAs connected to a PC. *Proceedings of ACM Conference on Computer Supported Cooperative Work (CSCW)*, pages 285–294, 1998.

[17] M. A. Nacenta, S. Sakurai, T. Yamaguchi, Y. Miki, Y. Itoh, Y. Kitamura, S. Subramanian, and C. Gutwin. E-conic: a perspective-aware interface for multi-display environments. *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 279–288, 2007.

[18] M. A. Nacenta, S. Sallam, B. Champoux, S. Subramanian, and C. Gutwin. Perspective cursor: perspective-based interaction for multi-display environments. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 289–298, 2006.

[19] M. A. Nacenta, D. Aliakseyeu, S. Subramanian, and C. Gutwin. A comparison of techniques for multi-display reaching. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 371–380, 2005.

[20] D. Olsen and T. Nielsen. Laser pointer interaction. *Proceedings of Conference on Human Factors in Computing Systems CHI)*, pages 17–22, 2001.

[21] J. K. Parker, R. L. Mandryk and K. M. Inkpen. TractorBeam: seamless integration of local and remote pointing for tabletop displays. *Proceedings of Graphics Interface (GI)*, pages 33–40, 2005.

[22] A. Pavlovych and W. Stuerzlinger. The tradeoff between spatial jitter and latency in pointing tasks. *Proceedings of the 1st ACM SIGCHI symposium on Engineering interactive computing systems (EICS)*, pages 187–196, 2009

[23] J. S. Pierce, A. S. Forsberg, M. J. Conway, S. Hong, R. C. Zeleznik, and M. R. Mine. Image plane interaction techniques in 3D immersive environments. *Proceedings of Inter-active 3D Graphics (I3D) 1997*, pages 39–43 1997.

[24] T. Prante, N. A. Streitz, and P. Tandler. Roomware: computers disappear and interaction evolves. *Computer 37*, pages 47–54, 2004.

[25] J. Rekimoto. A multiple device approach for supporting whiteboard-based interactions. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 344–351, 1998.

[26] J. Rekimoto and M. Saitoh. Augmented surfaces: a spatially continuous work space for hybrid computing environments. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 378–385, 1999.

[27] J. Rekimoto. Pick-and-Drop: a direct manipulation technique for multiple computer environments. *Proceedings of ACM Symposium on User Interface Software and Technology (UIST)*, pages 31–39, 1997.

[28] S. Sakurai, Y. Itoh, Y. Kitamura, M. A. Nacenta, T. Yamaguchi, S. Subramanian, and F. Kishino. A middleware for seamless use of multiple displays. *Proceedings of International Workshop on Design, Specification, and Verification of Interactive Systems (DSV-IS). LNCS*, vol. 5136, pages 252–266. Springer, Heidelberg, 2008.

[29] C. Schmandt and E. A. Hulteen. The intelligent voice-interactive interface. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 363–366, 1982.

[30] L. E. Sibert and R. J. K. Jacob. Evaluation of eye gaze interaction. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 281–288, 2000.

[31] R. W. Soukoreff and I. S. MacKenzie. Towards a standard for pointing device evaluation, perspectives on 27 years of Fitts' law research in HCI. *Int. J. Human-Computer Studies* 61, pages 751–789, 2004.

[32] W. Stuerzlinger and J. Oh. Laser pointers as collaborative pointing devices. *Proceedings of Graphics Interface (GI)*, pages 141–150, 2002.

[33] D. S. Tan, B. Meyers, and M. Czerwinski. WinCuts: manipulating arbitrary window regions for more effective use of screen space. *Extended Abstracts on Human Factors in Computing Systems (CHI)*, pages 1525–1528, 2004.

[34] R. J. Teather and W. Stuerzlinger. Assessing the Effects of Orientation and Device on (Constrained) 3D Movement Techniques. *Proceedings of IEEE Symposium on 3D User Interfaces (3DUI)*, pages 43–50, 2008.

[35] S. Voida, M. Podlaseck, R. Kjeldsen, and C. Pinhanez. A study on the manipulation of 2D objects in a projector/camera-based augmented reality environment. *Proceedings of Conference on Human Factors in Computing Systems CHI)*, pages 611–620, 2005.

[36] N. F. Yang, D. W. Jin, M. Zhang, C. H. Huang, and R. C. Wang. An extending Fitts' law for human upper limb performance evaluation. *Proceedings of Annual EMBS International Conference*, pages 1240–1243, 2001

[37] S. Zhai, C. Morimoto, and S. Ihde. Manual and gaze input cascaded (MAGIC) pointing. *Proceedings of Conference on Human Factors in Computing Systems (CHI)*, pages 246–253, 1999.