Ji-Young Oh*
Hong Hua
College of Optical Sciences
University of Arizona
Tucson, AZ 85721

Usability of Multi-Scale Interfaces for 3D Workbench Displays

Abstract

We consider that multi-scale visualization interfaces support users to view different levels of scales simultaneously and to understand large-scale, complex 3D information in 3D display environments. This article presents a user evaluation on three multi-scale interfaces on a 3D workbench display: focus + context (f + c), fixed f + c, and overview + detail (o + d). The interfaces differ in terms of (1) window arrangement and (2) positioning of detailed information relative to the user. Our goal is to identify the effect of these interface differences in large scale information visualization on a 3D workbench. To address the usability of the interfaces for a wide range of applications, we designed two tasks that differ by the level of information integration and cognitive demand. The evaluation results suggest that focus-based interfaces (i.e., the f + c and fixed f + c interfaces) are useful for tasks that require tight coupling between information layers and the o + d interface is useful for tasks performed in a densely populated information space. In terms of interface design on a 3D workbench, it is important to provide an up-close view of the current region of interest for fast scene navigation and an easy way to change viewing direction to see the 3D information from more comfortable directions. The detailed design guidelines based on the evaluation analysis are presented in this article.

1 Introduction

Multi-scale display interfaces have been well recognized for their crucial roles in navigating and understanding complex information spaces. The typical use of a multi-scale interface may be for visualizing information that is too small/large or too dense to be displayed in a single scale. A more profound impact of such interfaces lies in understanding the hierarchical structures of complex information spaces: a macro scale may help in understanding the overall structure of the information space, for example, a pattern or trend of events, and in planning tasks; a micro scale facilitates working on a local detail, such as inspection, local alignment, or simple observation. The field study conducted by Baudisch, Good, Bellotti, and Schraedley (2002) demonstrates that multi-scale interfaces are essential to a wide range of applications, from professional to entertainment, such as graphic design, mapping, electronic chip design, and gaming.

Numerous multi-scale interface techniques have been explored for visualizing and navigating 2D and 3D graphical information, many of which were suc-

*Correspondence to Ji-young.oh@autodesk.com.

Presence, *Vol. 17, No. 5, October 2008*, 415–440

cessfully adopted in commercial applications. For example, the panning and zooming technique and interfaces with an overview (Plaisant, Carr, and Shneiderman, 1995) are commonly utilized in 2D image manipulation and map navigation, such as Google Maps. A substantial number of empirical studies have investigated the benefits and problems of such interfaces for complex information visualization in 2D display environments (Gutwin and Fedak, 2004; Hornbaek, Bederson, and Plaisant, 2002; Bederson and Hollan, 1994; Baudisch et al., 2002; Reilly, Rodgers, Argue, Nunes, and Inkpen, 2006). The added dimensional complexity of 3D display environments, in particular those dealing with real-time user interaction with virtual and augmented environments, suggests potentially an even greater need for multi-scale interfaces. However, few empirical studies have been conducted to investigate the usability and effectiveness of multi-scale interfaces in 3D display environments where the extended dimensionality imposes unique challenges on interface design methods and usability concerns.

In this article, we present the empirical evaluations of multi-scale interface techniques in a 3D workbench environment that is built as a part of our SCAPE (Stereoscopic Collaboration in Augmented Projective Environment) infrastructure (Hua, Brown, and Gao, 2003; Hua, Brown, and Gao, 2004). With this work, we aim to understand how different multi-scale interfaces affect the mental integration of different levels of detail (LOD) of information in different tasks and produce interface design guidelines for multi-scale 3D information visualization on stereoscopic workbench environments.

We consider interfaces that simultaneously present multiple LODs of a complex information space, rather than using the sequential panning and zooming technique. Such interfaces allow concurrent interaction with different scales. In particular, the three interfaces to be tested were adapted from two of the most successful multi-scale interfaces in the 2D domain: *focus plus context* (f + c) visualization (Furnas, 1986; Carpendale and Montagnese, 2001) and *overview plus detail* (o + d) interface (Plaisant et al., 1995) into our 3D workbench environment. The layouts of the three interfaces are demonstrated in Figures 1a, 1b, and 1c, respectively.

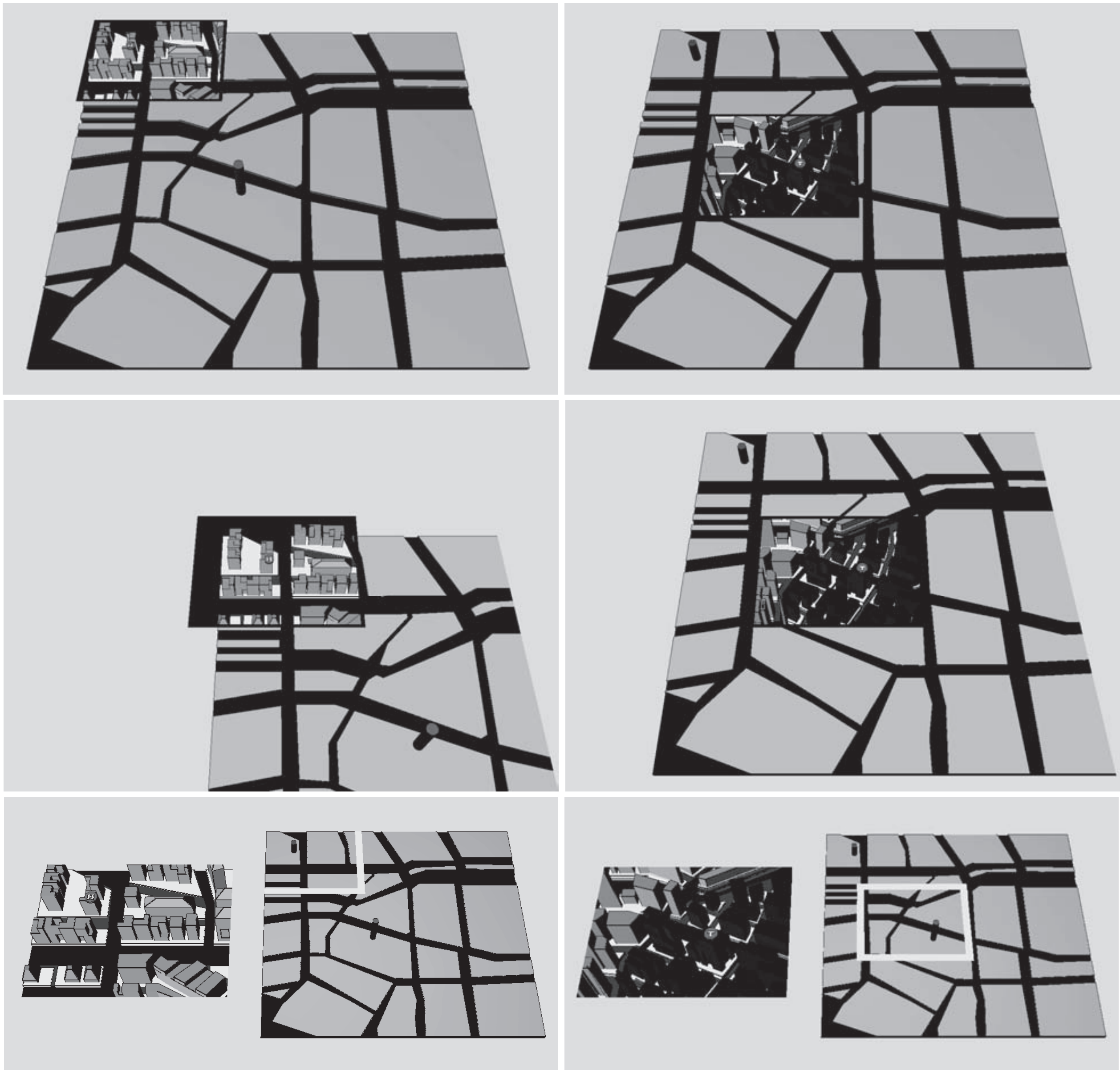
The f + c interface presents zoomable, detailed information through a small focus lens embedded in a larger context window, while the o + d interface provides the zoomable detail window separate from the overview window showing the entire information space. The two interfaces differ not only in terms of window arrangement, but also in terms of methods to position the focus or detail window relative to a viewer. More specifically, while in f + c, the focus window typically moves over a large context display, the detail window in o + d is in a fixed position and the scene appears to pan into the detail window. This is an important factor especially for a large 3D workbench environment where (1) the physical distance between a user and the scene significantly affects information readability, and (2) the relative motion between the user and the scene changes viewing direction in a 3D space and consequently affects the visibility of 3D information (e.g., an object may be occluded from one view direction but not from another direction). To account for this factor more directly, we added one more interface in our evaluation: *fixed focus plus context* (fixed f + c). The fixed f + c interface has the same window arrangement as the f + c; however, its focus window is fixed in the center and the context pans into the focus area as in the o + d interface. The fixed f + c interface shares similarity to the implementation of a focus plus context screen by Baudisch et al. (2002) in terms of focus window positioning.

2 Related Work

2.1 2D Multi-Scale Interfaces

Two-dimensional interfaces for multi-scale information visualization have been extensively studied in desktop environments (Gutwin and Fedak, 2004; Hornbaek et al., 2002; Bederson and Hollan, 1994), large-scale projection-based displays (Baudisch et al., 2002; Lee, Hudson, Summet, and Dietz, 2005), and handheld devices (Reilly et al., 2006; Baudisch and Rosenholtz, 2003).

The most basic multi-scale interface involves panning and zooming sequentially to obtain the desired view of



(a)
(b)
(c)

Figure 1. *Window arrangements and manipulation schemes of the three interfaces, left column at the starting point and right column at a new position of interest: (a) f + c, (b) fixed f + c, and (c) o + d.*

information (Bederson and Hollan, 1994; Hornbeak et al., 2002). On the other hand, the multi-window approaches (Plaisant et al., 1995; Baldonado, Woodruff, and Kuchinsky, 2000) allow simultaneous view of multiple levels of scale. The o + d interface is a type of multi-window approach, where a user progressively zooms and pans the document in a detail window, while an overview window displays the whole information space (Bederson and Hollan, 1994). The overview window typically contains an inset marker highlighting a user's current area of interest that is coordinated with the detail window. A number of empirical studies have been conducted for usability issues of o + d interfaces (Hornbaek et al., 2002; North and Shneiderman, 2000).

Hornbaek et al. compared an o + d interface with the panning and zooming interface (i.e., an interface without overview) for map browsing tasks. The study results suggest that the separate overview window divides users' attention, resulting in their recalling fewer map objects with the o + d interface than with the detail-only interface. However, the study shows that most users still preferred the interface with an overview. North and Shneiderman (2000) evaluated the effect of coordinate correlation between the detail and overview windows on task performance and user preference. The results show that users strongly prefer a tightly coupled o + d interface over a noncorrelated o + d interface or a detail-only interface.

Another multi-window approach is to embed the focus or detail area into a large context window, such that the window arrangement is traditionally considered as a type of Magic Lens metaphor (Bier, Stone, Pier, Buxton, and DeRose, 1993). Due to the overlaid layout of the windows, a zoomable focus display can result in considerable information discontinuity between the embedded focus and the context views with different scales applied. To provide a smooth transition between the focus and the context views, distortion-based visualization techniques such as fish-eye lenses have been extensively studied. Excellent discussion of these techniques can be found in Furnas (1986), Leung and Apperley (1994), and Carpendale and Montagnese (2001); and empirical studies of the techniques can be found in Gutwin and Fedak (2004) and Hornbaek and Frokjaer (2001). However, the distortion-based interface imposes a critical disadvantage in applications where accuracy of interaction is required. In order to address this problem, Baudisch et al. (2002) introduced an f + c screen which incorporates a desktop LCD monitor in a large-size, low-resolution projection display. The LCD monitor serves as a high-resolution focus display placed in the center of a surrounding low-resolution context display. The usability studies suggest that the f + c screen can improve a user's task performance by 20% over an o + d interface. However, the tasks in the studies were biased toward the f + c screen in the sense that the tasks required strong correlation between focus and context information. Additionally, a fixed magnification factor was set between the focus and the context windows, which eliminated "zooming" time using this particular interface.

2.2 3D Multi-Scale Interfaces

Various approaches to multi-scale interfaces have been explored for 3D scene navigation and information visualization. The "World in Miniature" (WIM) interface presented by Stoakley, Conway, and Pausch (1995) may be one of the earliest attempts to provide an o + d interface in a 3D Immersive Virtual Environments (IVE). The WIM interface offers a 3D overview of the whole world in a low LOD to assist navigation in an IVE without losing location awareness. The original WIM interface is limited to a single, low-LOD overview display. To overcome this limitation, scalable WIM interfaces have been proposed by LaViola, Feliz, Keefe, and Zeleznik (2001) and Wingrave, Haciahmetoglu, and Bowman (2006). Alternatively, Pierce and Pausch (2004) extended the basic WIM interface with hierarchical maps that determine visible landmarks to assist navigation in a large-scale virtual environment.

The Magic Lens type of multi-scale interfaces has been explored in 3D environments by Viega, Conway, Williams, and Pausch (1996). Schmalstieg and Schaufler (1999) introduced a life-sized Magic Lens directly into a 3D virtual environment to support sharing views. Mendez, Kalkofen, and Schmalstieg (2006) introduced a context-sensitive volumetric Magic Lens, where each virtual object stores its reaction to different types of Magic Lens operation so that different visual styles of multiple objects can be concurrently visualized through a single lens window.

Several efforts have been made to explore Magic Lens-type interaction schemes in Augmented Reality (AR). For instance, the work of metaDesk by Ullmer and Ishii (1997) used a handheld display as a Tangible Magic Lens to visualize the image of a selected area of interest on a 2D workbench display. More recently, Looser, Billinghurst, and Cockburn (2004) explored Magic Lens interaction schemes in their marker-based AR system.

Previously, we developed a SCAPE infrastructure for multi-scale information interaction in an Augmented Virtual Environment (Hua et al., 2003; Hua et al., 2004). SCAPE is composed of an immersive wall display and a workbench, where the workbench installed inside the room display operates as a WIM to the world in which a user is immersed. The SCAPE system further adopts various forms of Magic Lens metaphors to support multi-scale interfaces on the workbench and wall displays (Brown, Hua, and Gao, 2003; Brown and Hua, 2006; Oh and Hua, 2006).

To our best knowledge, only a few empirical studies investigate usability issues of 3D multi-scale interfaces. Furthermore, there is no interface design guideline that provides common ground across the existing interface implementations, which motivated our study presented in this paper.

3 Interfaces and Apparatus

Figure 1 demonstrates the three interfaces, f + c, fixed f + c, and o + d, to be evaluated in our study. Although these interfaces have been explored in the 2D domain to some extent, they have to be adapted in a sensible manner to accommodate user interaction in a 3D workbench environment, which imposes new challenges in interface design and raises additional usability concerns.

Due to the importance of accurate and continuous depth perception in a 3D environment, in both types of the focus-based interfaces (f + c and fixed f + c), we provide an undistorted focus view, rather than applying distortion-based visualization techniques (e.g., fish-eye lenses) that are typically adopted in 2D f + c interfaces (Carpendale and Montagnese, 2001). As a consequence, the focus-based interfaces provide tight coupling between the focus and context windows when the two views are displayed in the same scale factor. However, if the focus window is magnified significantly larger than the context, the discrepancy of the scale factors degrades the coupling between the two views. An additional problem resulting from magnification is the occlusion of context information underneath the focus window, which exacerbates information discontinuity.

In the o + d interface, the detail window is positioned at the center of the workbench, easily accessible to the user, which is similar to the fixed f + c interface. In contrast to the two focus-based interfaces, it provides a separate overview. Consequently, the detail window does not block the view to the overview and yields no information discontinuity with increasing magnifications. However, the o + d interface may demand a user to switch his or her attention between the two views to keep location awareness, which requires mentally integrating spatial information from the two separate windows.

A large workbench environment and the nature of 3D visualization lead to unique use conditions for the interfaces, which can significantly affect information readability and visibility. Firstly, common to any tabletop display, the table size is a critical problem to designing an interface that is reachable and viewable from a viewer's position. Secondly, since a 3D scene varies in height and orientation, some information may be occluded and thus impossible to read from a certain viewpoint, necessitating an interface mechanism to provide easy access to the occluded data. Thirdly, it is oftentimes difficult to incorporate symbolic information into the 3D visualization, due to the conflict with viewing directions and scene clutter. We conjecture that the two levels of detail between overview and detail, or context and focus, may help users to more easily understand the pattern of the information space. We attempt to address how the different arrangements of interfaces allow users to incorporate different levels of information.

The fixed f + c interface is introduced to account for the differences between f + c and o + d. In fixed f + c, the scene pans into the focus window that is fixed in the center of the workbench, making any part of the context scene equally accessible to the user. A compromise lies in the fact that some important part of the context may be cut off as the user pans the scene to a side.

Through the evaluation, we intend to identify how users cope with the differences among the three interfaces and how the different factors in the interfaces affect user performance. In particular, we intend to study the effect of windows arrangement and the information readability issues of focus windows and attempt to address the following questions:

1. How does the window arrangement of a multi-window interface (focus embedded or detail sepa-

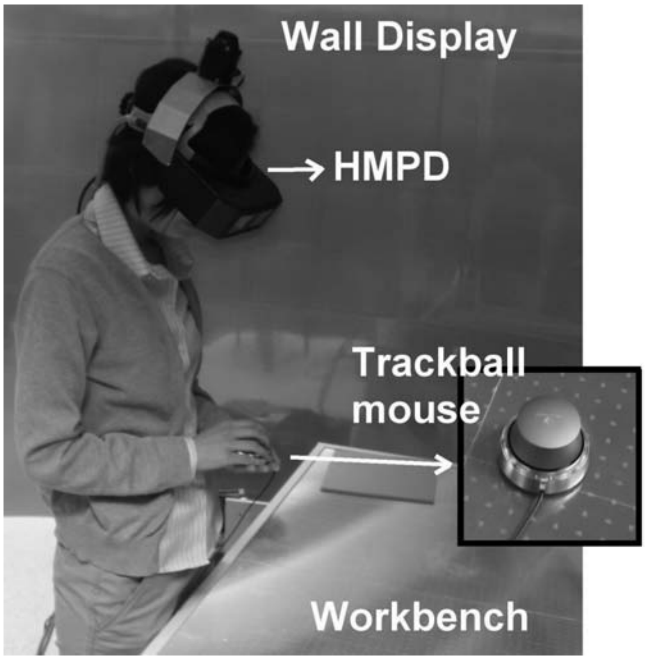


Figure 2. *Evaluation setup in the SCAPE system.*

rate) affect navigation activity when the task requires a tight coupling between the focus and context or overview and detail?

2. How do the different window manipulation schemes in f + c and fixed f + c affect information readability and visibility issues (e.g., occlusion and scale of objects)? In addition, how does a user cope with the lost context information at the edge of a display in fixed f + c as a result of panning the context instead of the focus?
3. How do the three interfaces affect task performance and user preference for tasks that require only loose coupling between the two levels of detail?
4. Will panning the scene toward the focus or detail windows in o + d and fixed f + c facilitate more dynamic 3D perception cues over the moveable focus window in f + c?

3.1 Apparatus

The evaluation of the 3D multi-scale interfaces described above will be performed in a 3D display environment called SCAPE (Figure 2). SCAPE is built upon a head-mounted projection display (HMPD) technology originally presented by Fisher (1996) and Kijima and Ojika (1997). It is composed of micro-displays, projection lenses, and beam splitters which are assembled in a helmet, and retro-reflective screens which are strategically placed in the environment. One of the important properties of a retro-reflective screen is that the light projected onto the surface is reflected right back toward its source. This property ensures that a pair of stereoscopic images is inherently separated and thus enables stereoscopic viewing capability. More in-depth discussion on HMPD technology can be found in Hua, Girardot, Gao, and Rolland (2000) and Hua, Gao, Biocca, and Rolland (2001).

The unique combination of the retro-reflection and projection properties of the HMPD technology allow us to build more flexible display environments than CAVE-like projection displays. For instance, any continuous surface that is coated with a retro-reflective material works as a display surface where virtual objects can be viewed through the helmet. The SCAPE infrastructure consists of a 6 × 3 ft interactive workbench and a 12 × 12 × 9 ft room-sized walk-through display environment that are coated with the retro-reflective material (Figure 2; Hua et al., 2004). In the SCAPE display paradigm, an exocentric view of a world is registered on the workbench, while a corresponding immersive scene is visualized in the surrounding room. Several auxiliary display widgets render views in intermediate scales bridging the levels of detail provided by the room and workbench displays. Providing that the focus of our user study is on windows arrangement of multi-scale interfaces on a 3D workbench, all the interactions in this article are performed on the workbench display in the SCAPE infrastructure.

3.2 Interface Setup

In the f + c interfaces, a large context window generally displays the entire information space in a low LOD with an embedded focus window displaying a subregion of the space in higher LOD (Figure 1a, b and Figure 3a, b). Therefore, in our interface setup, the whole 3D scene is scaled to fit in the vertical dimension of the workbench area. As a result, the dimension of the context window is approximately 100 × 86 cm, while the workbench measures 180 × 90 cm. The size of the

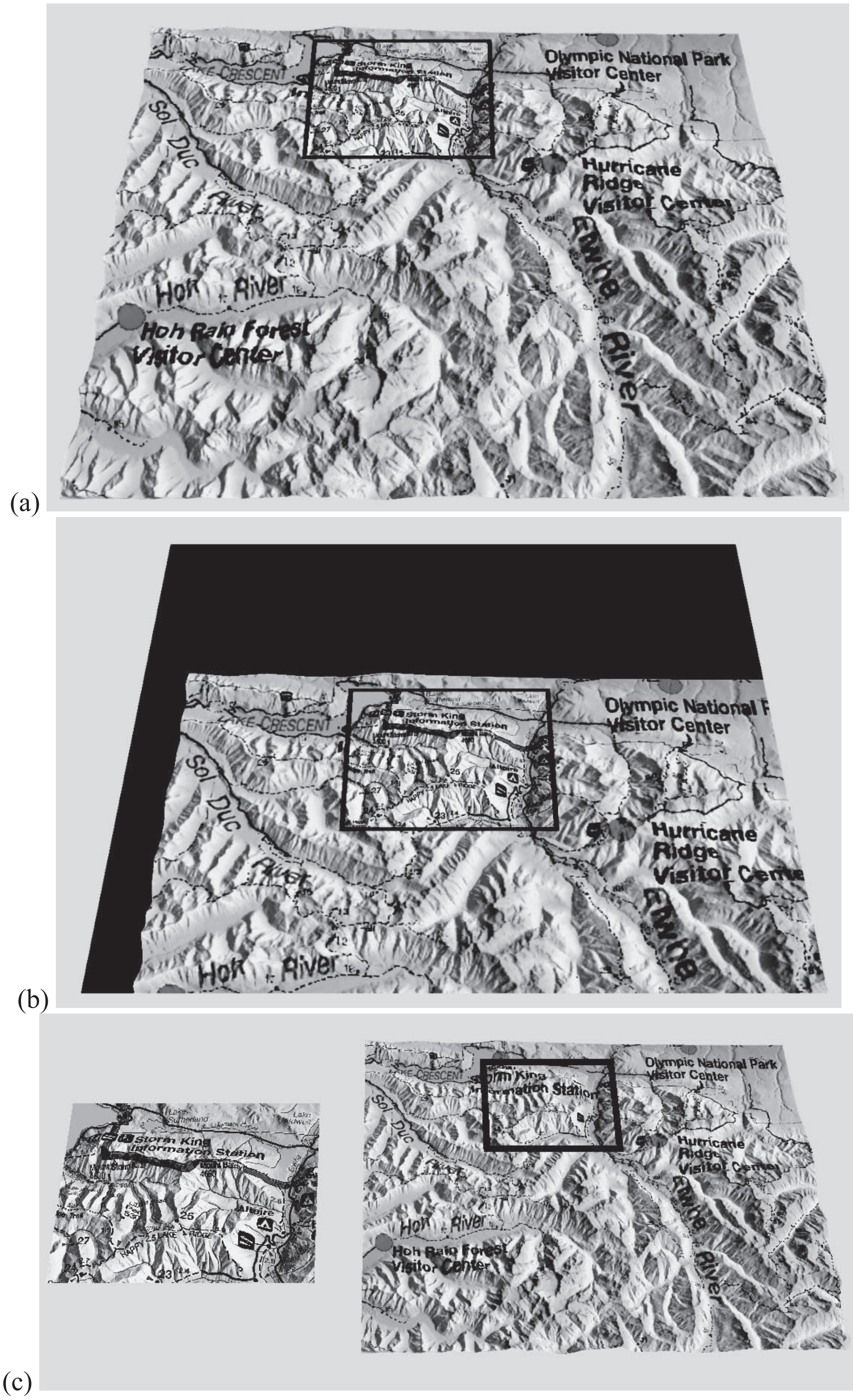


Figure 3. *The three interfaces in the map reading task based on Olympic National Park map overlaid with its terrain: (a) f + c, (b) fixed f + c, (c) o + d.*

focus window was based on two factors: (1) it should be smaller than the instantaneous field of view (FOV) of our HMPD display, so that a small portion of the context information is simultaneously visible when a user is browsing the focus window; (2) it should be big enough to accommodate a usable amount of highly detailed information. Providing that the projection area of our HMPD FOV on the workbench measures 52 × 39 cm for average users at a standing pose (assuming 70 cm above the workbench), we set the focus window to be 40 × 30 cm. The ratio of the focus window area to the context window is about 7.2.

In our evaluation, we strove to keep the amount of information presented to the users consistent across the interfaces, which requires that the detail window in o + d and the focus window in f + c and fixed f + c should be in the same dimension (40 × 30 cm) and the containing scene should be displayed in the same scale. The default scale of the scene in the detail or focus window is decided by the scale of the context window, which provides a tight coupling between the focus and context windows when no magnification is applied.

In the o + d interface, we have provided separate detail and overview windows placed side by side, as shown in Figures 1c and 3c. This arrangement was adopted from Baudisch et al. (2002). The dimensions of the overview window are based on the measurements of the projection area of the HMPD's FOV. The overview window should be small enough so that users can look at the whole scene in one quick glance, but large enough to accommodate the contents with enough resolution. The width of the overview window was thus set to be around 50 cm and its height varied around 40 ± 3 cm, depending on the scene's original aspect ratio. It is noteworthy that the dimension of HMPD's projection area is not fixed, but proportional to the distance between users' eyes and the workbench surface. If a user steps back a little farther away from the workbench, he or she will be able to see a larger projection area. The overview window is separated from the detail window by about 9 cm on the workbench.

Finally, the horizontal dimension of the context window is approximately equal to the total horizontal extent of the overview and detail window in the o + d interface. This configuration ensures that all three interfaces require a user equivalent level of physical "walking" in front of the workbench to access the entire information space.

It is also noteworthy that the control of all three interfaces is in 2D, that is, the windows motion is constrained to the horizontal surface of the workbench display. To provide 6 degrees of freedom (DOF) interfaces, we must choose a proper interface device and an interface window manipulation method from a number of possible alternatives. For example, the window can be manipulated via a go-go method (Bowman and Hodges, 1997) or a tangible prop. In fact, we already have developed the tangible Magic Lens (Brown and Hua, 2006; Oh and Hua, 2006), which can be considered as a 6 DOF version of the f + c interface in this paper. However, providing a 6 DOF adaptation of the o + d interface that has the consistent manipulation metaphor with the tangible Magic Lens is a challenging problem that needs another series of research. Furthermore, since there can be various ways to achieve the 6 DOF control for the three interfaces, this will add another dimension to our current evaluation design. To eliminate the interface implementation factors that are not inherent to the definition of interfaces, we chose the standard 2D control interface, using a trackball mouse as shown in Figure 2.

Consequent to the choice of 2D interface control, we used task datasets of which base planes are tightly registered on top of the workbench surface, such as city models and 3D terrains of national parks, rather than datasets that float above the workbench possibly like medical or molecular datasets. The map-like datasets are not only commonly tested in 2D interfaces, but also represent a wide range of possible applications in 3D workbench environments, such as urban design or geographic information systems. Finally, such types of datasets adequately present important usability issues unique to 3D visualizations, such as occlusion and 3D perception problems. Therefore, we expect our study results using the 3D models of cities and national parks will present a good contrast with the previous studies in 2D display interfaces and embrace a range of real-world application scenarios. Eventually, the study will provide us with a good foundation to design interfaces that can be used for different types of datasets with more extended dimensions.

4 Tasks and Measurements

We designed two types of tasks: path following and map reading. The path following task requires a user to use combined information from the two windows of different levels of detail (focus/detail and context/overview) to follow a shortest path and navigate from one point to the other point in a 3D city model (Figure 1). On the other hand, the map reading task requires a user to gather 3D information of two map features from the focus or detail view, whose locations are specified relative to the landmarks displayed in the respective context or overview window. The datasets we have used for the two types of tasks are imaginary city models and real national park maps, respectively.

4.1 Path Following Task

The main purpose of the path following task is to examine the effect of the windows arrangement in the three interfaces on applications where strong coupling between the focus and context information is required. The task is designed such that a user must combine information from both the focus and the context or the detail and the overview. The goal of the task is to navigate a shortest path from a specified starting point to a target point in a model of a fictitious city (Figure 1). In the macro-scale window (context/overview), major streets and two color-coded start and target points are displayed. In the micro-scale window (focus/detail), detailed streets, all the 3D buildings, and color-coded path markers that correspond to the start and end points in the context/overview are displayed. To find a shortest path, the user must refer to the context/overview for connections between major streets and the focus/detail view for minor branching streets, while simultaneously maintaining the current location of the focus/detail view relative to the endpoint that is visible in the context/overview. Once a user arrives at an endpoint, the user will have to read the alphabet letter on the marker visible in the focus/detail to designate to the experimenter that he or she completed the path. As soon as the marker is read, the experimenter shows the next target position to which the user must navigate from the current position. In one trial, a user pursues four consecutive target markers.

Figure 1 demonstrates the task in the three interfaces. The pictures in the left column show the focus/detail window on the start point, and those in the right column show the focus/detail window on the target point. In this task, we log time-stamped user interactions, containing focus/detail window location and user head motion. From the collected data, we produce the following measurements:

1. Task time: The measurement demonstrates the overall performance of each interface.
2. Movement distance: Taking the center position of the focus/detail window, we measure the distance of the path that a user has taken, starting from one point to the current target point. The distance designates the efficiency of the paths that are taken.
3. Movement speed: It is measured by using the movement distance and task time. Although this measurement appears to be redundant with the above two measurements, it can suggest how efficiently a user is able to view and integrate the multi-scale views.
4. User attention categorization: Due to the absence of eye-tracking in our current displays, we used head motion patterns recorded by the head tracker to categorize how users managed their attention to conduct the task. The detailed process of producing quantitative results is described in Appendix A.

It is noteworthy that we do not consider the zooming factor as a measurement for this task. The default scale factor set for the focus/detail window is generally sufficient for recognizing a path and reading the marker letter, thus we do not expect users to zoom in or out during the task, even though zooming is enabled.

4.2 Map Reading Task

The main purpose of the map reading task is to examine the effect of windows arrangement in the three interfaces for applications where only loose correlation between the focus and context information is required. Through this task, we are interested in testing the following factors: (1) a user's ability to read the focus/detail while simultaneously maintaining a contextual understanding of his or her location, and (2) effect of different positioning methods of the focus/detail window on 3D perception. We use a national park map overlaid on a 3D terrain, as shown in Figure 3. In many information visualization systems, symbolic information such as labels or icons along with geometric objects is indispensable to conduct tasks related to information gathering and problem solving. We anticipate that the symbolic information is important to evaluate the interfaces that incorporate a mixture of various types of information.

For each map browsing task, users are asked to find two specific places (e.g., mountains or rivers) in a national park displayed in the focus/detail, whose locations are specified by the surrounding landmarks displayed in overview/context (e.g., Obstruction Peak in the area east of Elwha River; here Obstruction Peak is visualized in the focus/detail view and the Elwha River is a landmark in the context/overview window). Once the two specific map features are identified, the user has to answer the task questions by comparing 3D shapes of the two places, such as slope or shape of ridges. For each map, we have devised three sets of questions, each composed of five map browsing questions; different question sets are used for the three interfaces. The question types can be largely categorized into one-area questions and two-area questions. In one-area questions, users look for two features from a single neighboring area, whereas in the two-area questions, users search for two features from separate areas. One-area questions are somewhat easier than two-area questions, simply because the searching area is smaller. Sample questions are given in Appendix B.

Similar to the path following task, we log time-stamped user interactions, containing focus/detail window location and user head motion. From the log data, we produce measurements such as task time, scale factor, and head motion pattern. Since we are further interested in finding out how focus window placement affects information readability and thus affects user interaction with an interface, the scale factor is logged as an important measurement. Different from the path following task, however, in this task we mainly focus on observation of users' behavior and opinions on each of the interfaces, as the process of map reading requires more cognitively "advanced" operations than following a path.

Two national parks, Glacier National Park and Olympic National Park, were chosen firstly because the distribution of the map features (mountains, rivers) is relatively uniform across the map, and secondly because the map and terrain datasets are readily available. The national park map images with symbolic text were downloaded from the corresponding national park website http://www.nps.gov/. The maps were then combined with the color-coded height profiles to facilitate understanding height information of the park terrains. The resulting images were used as detail maps. The overview maps were produced by removing all the text labels in the map, except for a handful of significant landmarks such as visitor centers and large rivers or lakes that divide the area of the national parks. The detail and overview maps were then overlaid on top of terrain datasets. The terrain of Glacier National Park was downloaded from http://nris.state.mt.us/nsdi/nris/el10/dems.html and Olympic National Park was from http://duff.geology.washington.edu/data/raster/tenmeter/byquad/. The terrain datasets were resampled to produce high-resolution (4097×4097 pixels) and low-resolution (513×513 pixels) terrain datasets, each overlaid with the detail (focus) and overview (context) map images, respectively. The coverage of Glacier National Park is 54 km $\times$ 43 km and Olympic National Park is 52 km $\times$ 39 km.

5 Evaluation Design

5.1 Procedure

The test was a (three interfaces) $\times$ (two tasks) factorial within-subject test. The test was divided into two

Table 1. *Evaluation Sessions*

	Session #1	Session #2
Path following	Block 1 (*city A*): 3 interfaces × 3 trials × 4 paths	Block 3 (*city A*): 3 interfaces × 3 trials × 4 paths
	Block 2 (*city B*): 3 interfaces × 3 trials × 4 paths	Block 4 (*city B*): 3 interfaces × 3 trials × 4 paths
Map reading	Glacier National Park: 3 interfaces × 5 map browsing tasks	Olympic National Park: 3 interfaces × 5 map browsing tasks

test sessions performed on two separate days, with each session taking about 1.5 hr. In each session, a user first conducted two blocks of path following tasks in two different imaginary cities and then one block of map reading tasks.

In the first session, a user was introduced to the SCAPE system after signing a consent form and filling out a questionnaire regarding demographic information. Then the experimenter explained the three interfaces and the task procedure to the participant. The participant was given some time (~5 min) to practice the use of the trackball mouse. One trial was randomly chosen for the practice session.

In each block of the path following task, the user conducted three trials of navigation tasks in a fictitious city, using each of the three interfaces. Each trial of the navigation task consisted of four consecutive paths specified by markers (as explained in Section 4.1). The order of interfaces and trials were randomized for each user to counterbalance the learning effect. Users were allowed to take a short break between the trials. Once finished with the task, the user discussed the usability issues with the experimenter.

Following two blocks of the path following task, the user performed a map reading task. First, a paper map was given to the user to familiarize himself or herself with the names of features in the national park, as well as map labeling conventions (e.g., blue letters for a label of a river). Once the user felt comfortable in reading map features, the experimenter explained the further test procedure. For each interface, a user carried out a set of map browsing tasks consisting of five different types of questions, as described in Section 4.2. On starting a task, the user himself or herself was responsible for reading out the map browsing task questions. This was to make sure the user understood the task question and to avoid miscommunication due to the unique national park feature names. During the browsing task, the experimenter reminded the user of the task question as needed to prevent him or her from turning away from the task to read the map question again. On finishing the task, the user filled out a questionnaire form and discussed usability issues.

The second session was the same as the first session, except that there was no introductory training. Furthermore, for the map reading test, the user performed the task using the Glacier National Park Map in the first session and then Olympic National Park in the second session. In this way, the potentially significant learning effect was limited to the Glacier National Park map task. The evaluation sessions are summarized in Table 1.

5.2 Subjects

Twelve participants (9 males, 3 females, age 19–40) from the University of Arizona were recruited. None of the participants had a prior experience with the SCAPE system. None of the participants was familiar with either of the two national park maps.

6 Path Following Task Results

6.1 Task Time

Table 2 and Figure 4 summarize the average time (s) taken to navigate the city from one point to the other for each trial block. The decrease in average time in the four blocks suggests a learning effect ($F_{11,3} = 43.86$, $p \approx 0$). The Tukey-Kramer post-hoc test shows that session 1 (block #1, 2) and session 2 (block #3, 4)

Table 2. *Average Task Time (s)*

		f + c		fixed f + c		o + d		
		Avg	*SD*	Avg	*SD*	Avg	*SD*	ANOVA test
S1	B1	16.00	7.15	15.05	8.21	18.47	7.83	$F_{11,2} = 2.16, p \approx .1387$
	B2	14.46	4.35	13.06	4.30	18.04	7.59	$F_{11,2} = 16.30, p < .001$
S2	B3	11.86	4.86	9.77	3.31	12.45	4.80	$F_{11,2} = 13.87, p < .001$
	B4	11.74	4.26	10.24	4.34	12.50	5.94	$F_{11,2} = 11.02, p < .001$
All		13.51	5.57	12.03	5.78	15.37	7.24	$F_{11,2} = 19.96, p < .001$

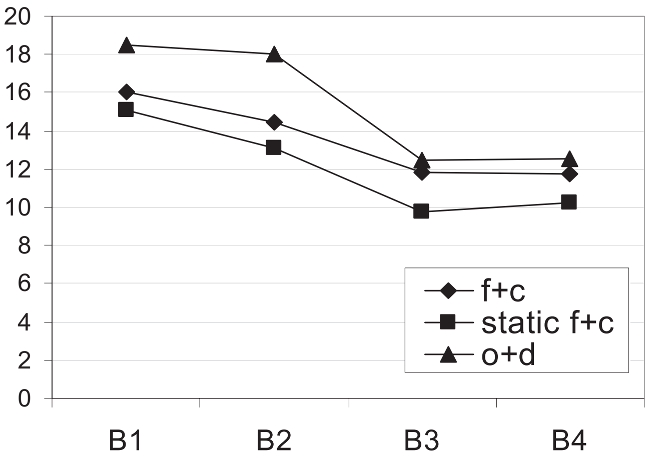


Figure 4. *Average task time (s) by trial block.*

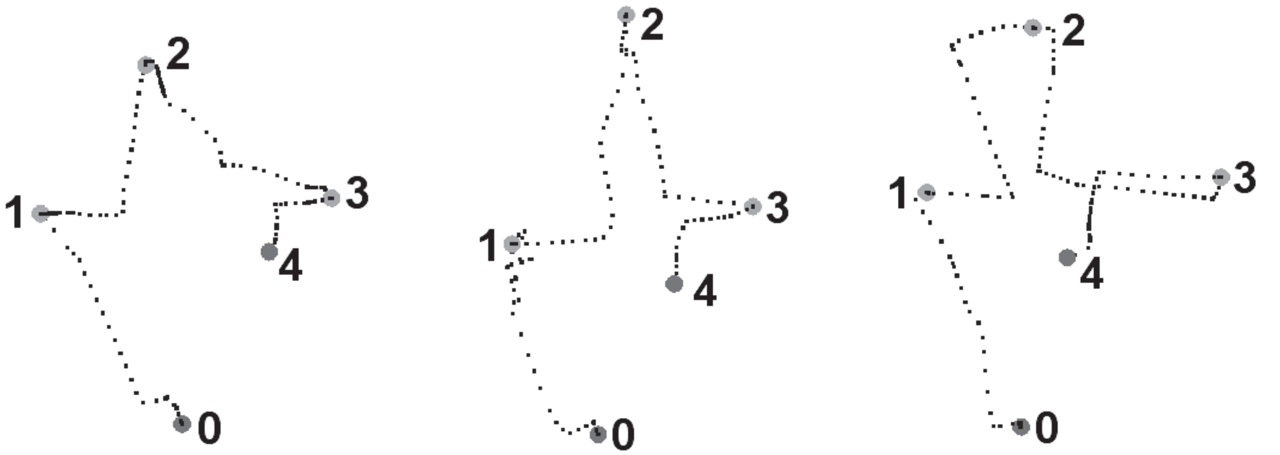
are different. It is noteworthy that users navigated the *City A* in block #1 and *City B* in block #2, both of which were performed in the first session. In the second session, users navigated different paths in the two previously-seen city models in block #3 (*City A*) and 4 (*City B*). Therefore, users were familiar with the city models in the second session. A repeated ANOVA test results in significant difference among the three interfaces ($F_{11,2} = 19.96$, $p < .001$). The Tukey-Kramer test shows that all three interfaces are different from each other.

The task time from the fixed f + c interface is 22% less than the o + d, and the f + c took 12% less than o + d. This result suggests that the embedded context awareness and centered position of the focus window in fixed f + c improve task performance over f + c and o + d.

6.2 Movement Distance and Average Speed

Based on the log data, we projected the trackball mouse paths onto the scale of the cities. The trackball mouse controls the center location of the focus/detail window. Figure 5 shows examples of the mouse motion projected onto *City B* by one user during a trial. A trial consists of four consecutive paths. In Figure 5, the start point is labeled as 0. The large dots labeled 1 through 4 correspond to the center of focus/detail window at the moment when a user confirms to the experimenter that he or she arrived at the target point. A user can see only two points at a time—current and next target point to navigate. A user takes a short break after finishing a path to target #4. The small gray dots indicate the trail of the trackball mouse between targets, sampled at every 30 ms. Since an area of interest (focus/detail window) occupies a certain amount of area and is not necessarily centered with the actual target point when completing a path, the locations of the targets in Figure 5 do not exactly match to each other among the three interfaces. Figure 5 demonstrates that the user has taken more efficient paths in f + c and fixed f + c than using o + d.

We have further compared total movement distance and movement speed to analyze how users performed the task with different interfaces. Figure 6a summarizes the total movement distance (cm) that users traveled from one point to the next target. A repeated ANOVA test results in significant difference among the three interfaces ($F_{11,2} = 9.54$, $p < .01$). The Tukey-Kramer multiple comparison test shows that movement distances using o + d are longer than the distances from



(a) (b) (c)

Figure 5. *User mouse motion projected onto a city model B in a trial composed of four targets. The starting and target positions are labeled numerically from 0 through 4 by their orders. The paths in interface: (a) f + c, (b) fixed f + c, (c) o + d.*

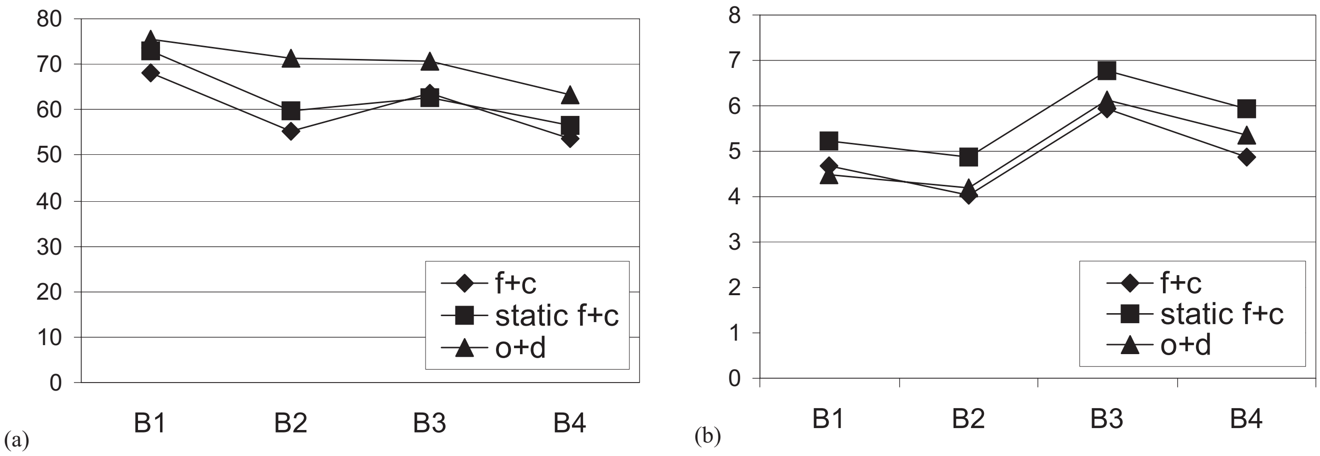


Figure 6. *The window movement pattern by trial block and interface: (a) movement distance (cm), and (b) average movement speed (cm/s).*

f + c and fixed f + c. This implies that users had difficulty in integrating information from the two separate interface windows using o + d. There is no difference suggested between f + c and fixed f + c.

Figure 6b shows the average movement speed (cm/s) by trial block. Overall, users navigated faster in the fixed f + c than both o + d and f + c ($F_{11,2} = 13.72$, $p < .001$). We conjecture that, through the center-positioned focus window of the fixed f + c, users were able to see the streets behind buildings more readily, while maintaining a relationship with the context easily. Interestingly, although users traveled short distances using f + c, similar to the fixed f + c, navigation speed using f + c was similar to the o + d interface. From the observation, users in f + c often had to physically bend over to see the occluded streets behind the buildings, which in turn slowed down the navigation speed. In o + d, users indeed had to take some time to coordinate the two separate views while looking back and forth between the two views.

6.3 User Attention Categorization

We categorized head motion pattern into three attention types: a user is viewing primarily the focus/detail window (*IN*), looking completely away from the focus/detail to attend the context/overview (*OUT*), or trying to look at the both of levels of detail windows at the same time (*BOTH*). Due to the absence of eye-tracking, we utilized the widely accepted eye-motion range (Melzer and Moffit, 1996) and foveal vision angle to set up the criteria to categorize users' head motion into the above three attention types. For a detailed description, please see Appendix A.

Figure 7 demonstrates the categorization of a specific user's attention during the task using the three interfaces. The pictures in the left column show the positions of the user gaze center relative to the interface windows on the workbench. The pictures in the right column show the mouse path through the city model. The gray boxes in the left column in Figure 7 represent the interface windows. More specifically, Figure 7a shows the trail of the moving focus window in the scene. The gray boxes in Figure 7c outline the detail window on the left side and the overview window on the right side. The dots in both columns are color-coded and symbol-coded by the category of the user's attention. The + signs represent IN, the × signs stand for OUT, and the • stands for BOTH. (These symbols are blue, red, and green, respectively, in the original.) During the tests, users generally stepped back and forth relative to the workbench to gain the proper viewing area and the best readability of the information. That is, users tended to step back to see a larger area to figure out overall direction toward the target marker, and they tended to bend toward the scene if they needed to look at small streets between buildings. Due to these head-to-table distance changes, there is some overlapping between *IN* and *BOTH* and between *BOTH* and *OUT*.

We calculated the percentage of time that a user spent viewing the specific windows in each interface relative to the total trial time. The results are shown in Figure 8. There was no significant difference among the interfaces for *IN* status ($F_{11,2} = 0.13$, $p \approx .878$). Significant differences were found in *OUT* ($F_{11,2} = 79.54$, $p \approx .00$) and *BOTH* ($F_{11,2} = 25.88$, $p \approx .00$). For *OUT* and *BOTH*, the o + d interface was different from the focus-based interfaces.

The results show that, using the focus-based interfaces, users mostly kept the view to the focus-window either in full attention (*IN*) or in partial view (*BOTH*). However, in o + d, users had to look completely away from the detail window to attend the overview. The result, in turn, directly demonstrates that the users' attention is divided by the two views while using the o + d interface. As it can be seen from the sample graph in Figure 7c, users tend *not* to pan the detail scene when they attend the overview window. Therefore, this may account for the significant amount of time spent viewing the overview in o + d, which may have partially contributed to the slower task time using this interface.

6.4 User Preference

When users were asked to choose the most favorable interface, a majority of users preferred f + c or fixed f + c over o + d. More precisely, out of 12 users, four users preferred f + c, three users preferred fixed f + c, four users preferred both f + c and fixed f + c equally, and only one user preferred o + d.

Many users mentioned they did not feel much of a difference between f + c and fixed f + c in terms of task performance because they have the same windows arrangement. There were some mixed opinions on fixed f + c, in that some users liked the interface because they did not have to move their heads when manipulating the focus window. Some users who did not like the fixed f + c stated that panning the 3D scene into the focus area was confusing for them. Several users mentioned the fixed f + c cut off the context area along with the next marker location, requiring them to make an extra motion to look for it.

Regarding f + c, a few users mentioned that they felt that it provided a narrower FOV, simply because the users had to move the focus window further away when looking at more distant target points, which, in turn, resulted in a smaller projection area of the focus window. Compared to fixed f + c, users liked the fact that they had the view to the whole context window all the time.

The major complaint on o + d was that users had to look back and forth between the two views. It required

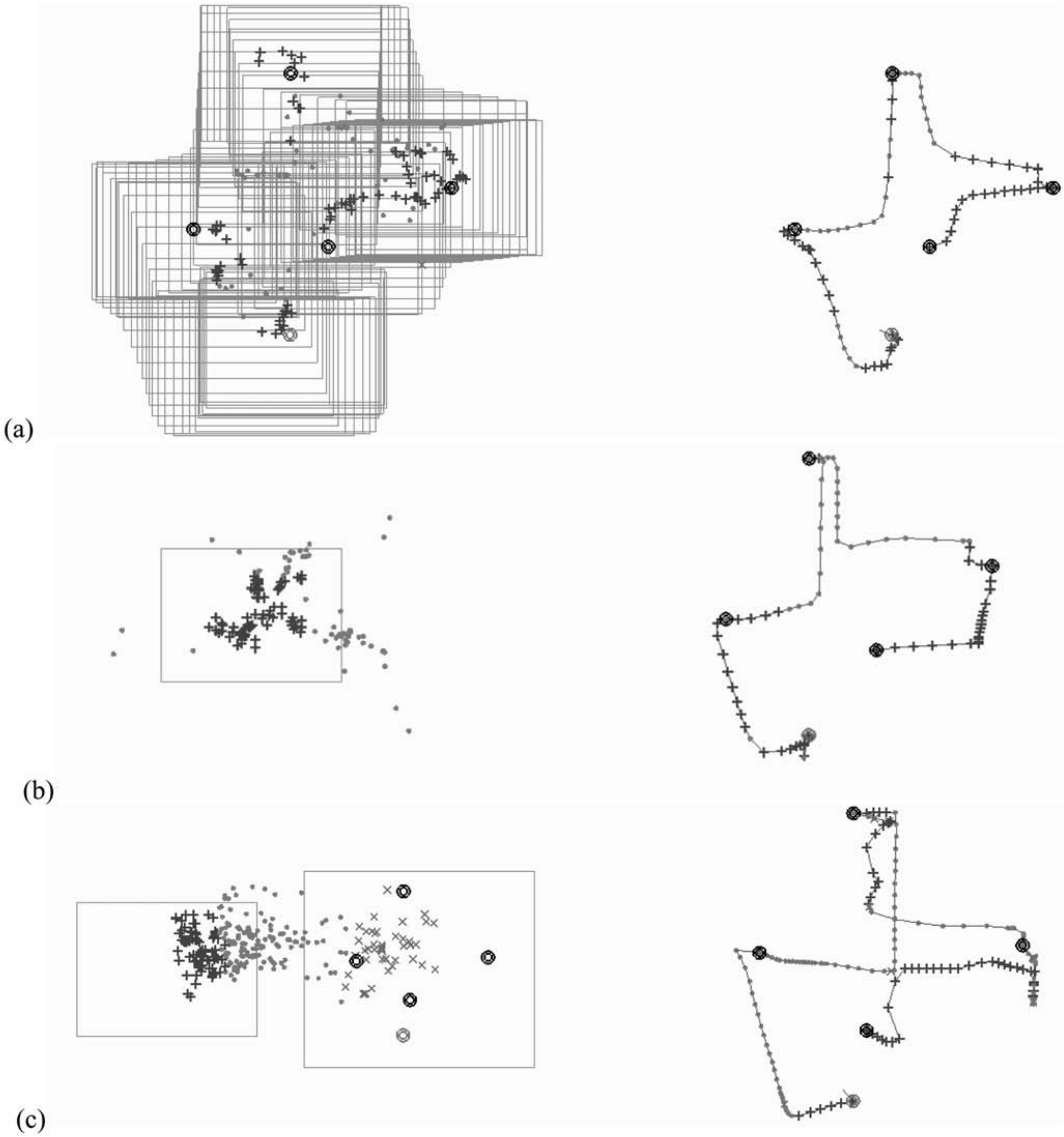


Figure 7. *The attention categorization results in each interface: (a) f + c, (b) fixed f + c, (c) o + d. Pictures in the left column are the user's gaze center relative to the interface windows projected onto the scale of the workbench. Pictures in the right column show the mouse paths projected onto the scale of the city model. All the dots are symbol-coded according to attention categorization: "+" for IN, "x" for OUT, and "•" for BOTH. The double circles are where the user must identify a marker. Gray boxes represent interface windows, and the double circles stand for markers defining the paths (purple in the original for the starting point).*

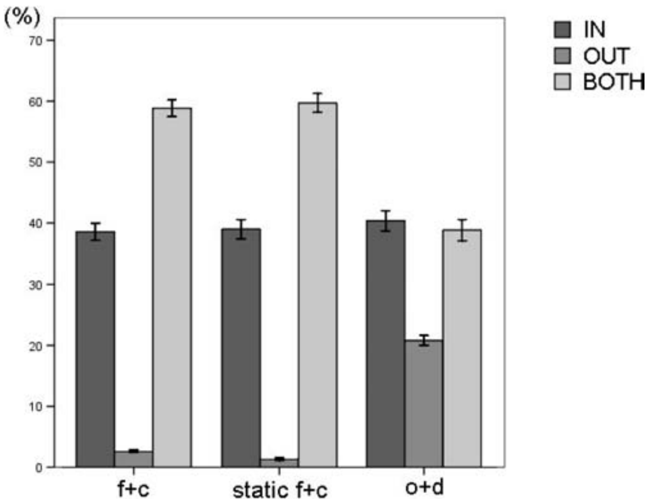


Figure 8. *Percentage of task time for each attention category.*

combining two views mentally to understand which main road shown in the overview branched to small roads in the detailed view. However, people mentioned that it was easier to see the location of a target point with a quick glance into the overview, which provides an unobstructed view of the entire scene, along with the user's current location.

7 Map Reading Task Results

7.1 Task Time

Table 3 summarizes the average task time by national park and question types. The browsing task required individual searching skills in an information-rich environment, which can vary widely by participants. Compared with the path following task, this task involves problem solving which demands a higher level of cognitive skill. Before the experiments, we had expected task times to be longer with the f + c than with fixed f + c or o + d, since f + c would require extra zooming time to read the labels far away. However, the results do not suggest such an effect.

7.2 Scale Factor

Figure 9 plots the average scale factors for each park map by interface. Interestingly, the scale factors from the tasks using f + c and fixed f + c are significantly larger than o + d in both park maps ($F_{11,2} = 7.36$, $p < .01$). It is not surprising that f + c requires a larger scale factor, since the focus window moves on the large extent of the context area, sometimes farther away from a user. However, we expected users would not scale up the scene using fixed f + c, as the far away features move *into* the center of the workbench, providing equal readability across any part of a map, just like o + d. Furthermore, in f + c and fixed f + c, a larger zooming factor results in the occlusion of more context information under the focus lens. On the other hand, in o + d, context information remains unoccluded, regardless of magnification. Therefore, we assumed that users would hesitate to use higher magnification when using the f + c interfaces, but this was not the case.

From our observation, users seemed to build a searching strategy based on the terrain information that is within their instantaneous field of view. For example, if a task requires finding a mountain, the user moves focus/detail along the ridges and peaks rather than browsing on flat terrain. The f + c and fixed f + c provide surrounding context regardless of zooming action, and thus users were able to utilize the context information for searching at all times. However, in o + d, as users magnified the scene, the detail covers a smaller area, hampering users' ability to predict the pattern of surrounding areas without referring back to the overview window. Therefore, users tended to compromise the scale factor in o + d to cover more of the search area, rather than engaging in the more time-consuming (and laborious) task of referring back to the overview window.

7.3 User Attention Categorization

Figure 10 shows an example of the user attention categorization and the mouse motion paths during the map reading task. The dots are color-coded in the original and symbol-coded, as explained in Section 6.1.3. Similar to the patterns observed from the path following task, users rarely moved the focus window completely away from their views. In o + d interface, the overview window was also used to navigate the scene. In contrast, during the path following task, users rarely panned the

Table 3. *Task Time by Map and Question Type (s)*

	f+c		fixed f+c		o+d	
	Avg	*SD*	Avg	*SD*	Avg	*SD*
Glacier, 1-area	64.81	45.90	63.46	60.87	65.47	49.39
Glacier, 2-area	73.37	33.33	66.66	44.73	73.95	45.56
Olympic, 1-area	39.10	27.17	32.53	12.53	40.59	32.32
Olympic, 2-area	54.82	27.64	63.87	41.77	68.33	39.48

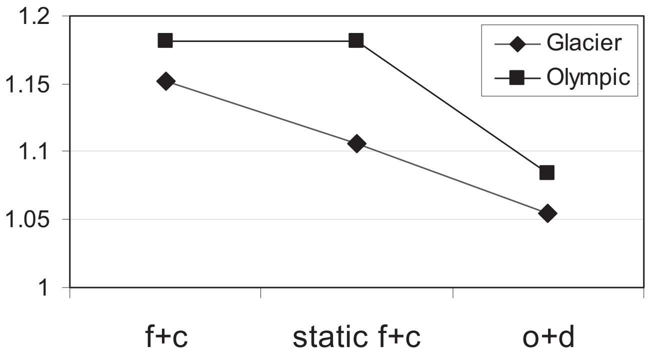


Figure 9. *Average scale factor.*

scene when they were viewing the overview window. This demonstrates the differences of our two tasks, in that the map reading task requires only loose integration of the two windows, and thus the overview was sufficient to navigate toward the approximate target region.

Figure 11 shows the different portion of attention categories by the interfaces and the maps. Consistent with the path following task, there was no significant difference in *IN* (Glacier: $F_{11,2} = 2.73$, $p \approx .09$, Olympic: $F_{11,2} = 0.94$, $p \approx .41$), but significant difference in *OUT* (Glacier: $F_{11,2} = 50.96$, $p \approx .0$, Olympic: $F_{11,2} = 149.99$, $p \approx .0$), and significant difference in *BOTH* (Glacier: $F_{11,2} = 25.95$, $p \approx .0$, Olympic: $F_{11,2} = 10.09$, $p < .001$). For status *OUT* and *BOTH*, the o + d interface was different from the focus-based interfaces.

The significant proportion of time spent in the *BOTH* status in the focus-based interfaces suggests that users were able to integrate the two different levels of information a lot easier in the focus-based interfaces than in the o + d interface.

7.4 Effect of Task Difference on User's Attention to the Overview Window

In order to find out how the task differences affected users' performance, we compared users' head motion patterns between the path following and the map reading tasks while using the o + d interface. Based on the process used in the previous sections (Section 6.3 and Section 7.4), we calculated the percentage of time spent viewing the overview window, the frequency (times per minute) of users' visits to the overview window, and the duration of overview attention.

Figure 12 illustrates the task difference and the user strategy according to the task. On average, users spent 20% of their time looking at the overview during the path following tasks, and 15% of their time during the map reading task. Users turned their heads to look at the overview at an average frequency of 7.48 times per minute in the path following task, and 2.52 times per minute in the map reading task. In path following, users spent 1.61 s in the overview per visit and 3.63 s per visit in the map reading task. Since we are comparing two different tasks, statistical testing is not appropriate.

The results demonstrate the difference between the two tasks in terms of the levels of demand for mental integration of the two views. The large differences in terms of the frequency of overview visits and average time spent per visit demonstrate that (1) users in the path following task made quick and frequent glances into the overview, and (2) on the contrary, in the map reading task, users did not look at the overview very often but spent a longer time per visit, mostly reading the landmarks in the overview. The longer overview visit time in the map reading task shows that the overview in

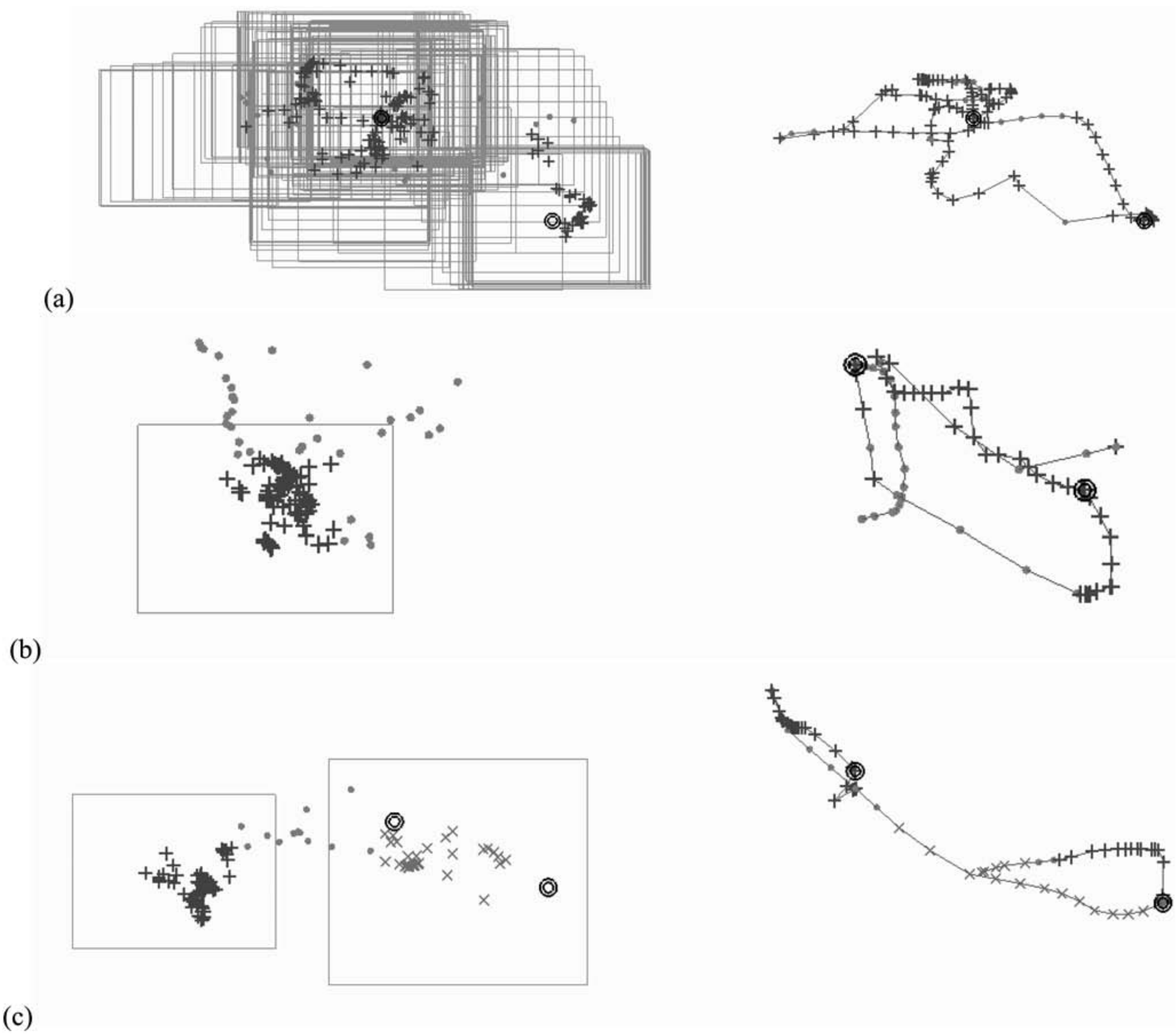


Figure 10. *Attention category during a map reading task: (a) f + c, (b) fixed f + c, (c) o + d. Pictures in the left column are the user's gaze center relative to the interface windows projected onto the scale of the workbench. Pictures in the right column show the mouse paths projected onto the scale of the national park map. All the dots are symbol-coded according to attention categorization: "+" for IN, "x" for OUT, and "•" for BOTH. The double circles are where the interest points were located by the user.*

the map reading contained more information than in the path following task.

7.5 User Preference

For the map reading task, most users preferred o + d (seven users); only two users preferred f + c, one user preferred fixed f + c, and one user stated no preference.

We believe the o + d interface was preferred since the task does not require users to keep *exact* correlation between the two views, but only requires users to keep track of a rough area designated by the landmarks in the context/overview. Accordingly, many users mentioned

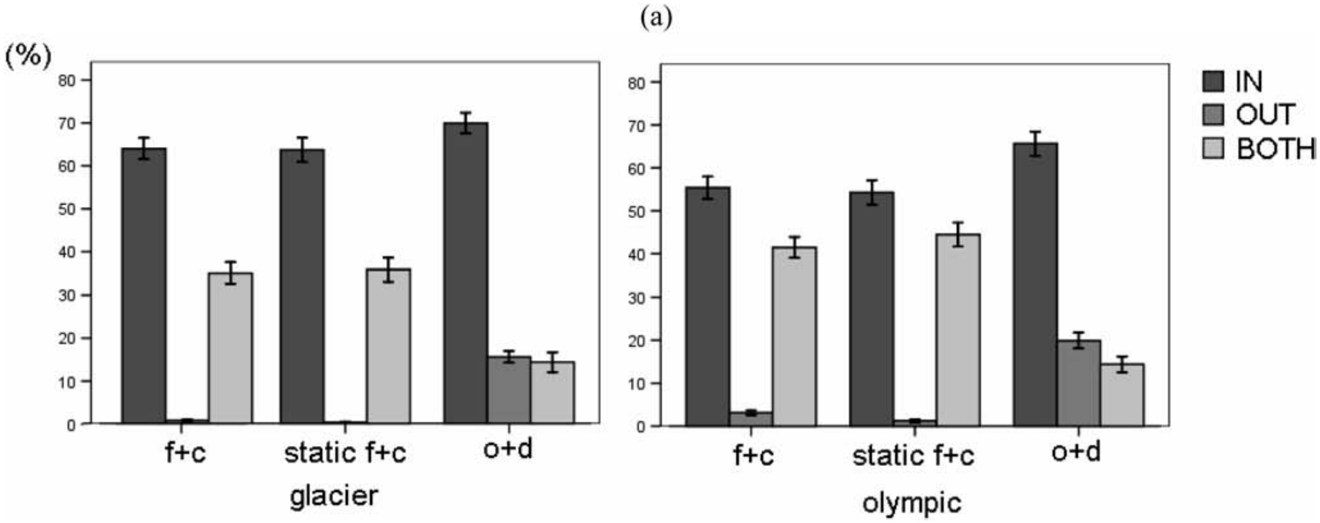


Figure 11. *Percentage of task time for each attention category by map.*

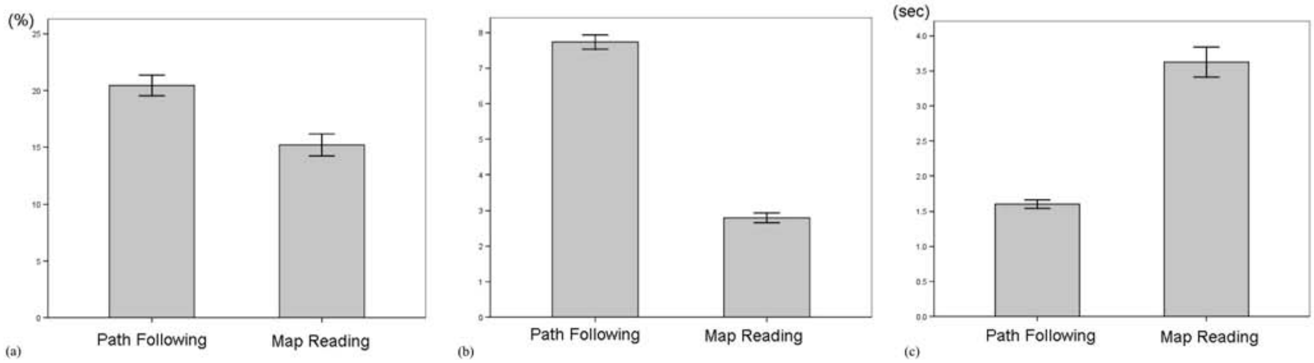


Figure 12. *Comparison of overview visits in o + d by task: (a) Percentage of time spent viewing overview to total task time, (b) number of overview visits per minute, and (c) time (s) spent per visit. Error bars show standard error of the mean,* SD/*SQRT*(N).

that the overview window in o + d provided a quick view to the overall scene, helping them to orient and decide the searching strategy. In addition, we conjecture that the text labels of the landmarks were easier and faster to read from the relatively small overview window rather than the large context window. Furthermore, the box marker indicated the position and size of the area of interest provided by the corresponding detail view. However, some users mentioned that they tended to drift away their focus area from the designated area more often when using o + d than f + c and fixed f + c.

As for f + c and fixed f + c, some users commented that the focus window partially blocks the view of the context area. Especially in fixed f + c, part of the context can be clipped off by the side of the workbench, providing even less area of context in effect. When zoomed in, some users feel that they lose the location awareness due to the mismatch of scales between the context and the focus. Users who liked f + c and fixed f + c thought that the interfaces provided better utilization of the field of view, as the context area can be seen simultaneously in the periphery, as opposed to the o + d, which requires looking away from the detail to see the overview.

Common to all interfaces, users complained of the neck fatigue induced by looking down onto the workbench for extended periods of time (~20 min). Some users also commented that it was problematic to understand shapes of 3D features when they looked straight downward.

8 Discussion

8.1 Usability in Terms of Interface Design

From the two different types of tasks—path following and map reading—we attempted to collect balanced information on usability of the three interfaces. The path following task emphasizes context awareness and ability of information correlation; the map reading task highlights readability of the inset window, while still requiring a loose degree of context awareness.

In the path following task, where a tight coupling between the focus/detail and context/overview is required, not only the windows arrangement but also the focus/detail window positioning method relative to the user affected the task performance.

The focus-based interfaces are more efficient for such types of tasks than the o + d interface. Users completed the task using fixed f + c 22% faster than o + d and 11% faster using f + c than o + d. The superior performance of the focus-based interfaces is due to their embedded arrangements of the windows, which provides better context awareness and requires less mental effort to correlate information between the windows than o + d.

Furthermore, in fixed f + c, the fixed position of the focus window with the panning scheme to move the scene toward a user afforded an up-close view into the focus area, allowing users to quickly look at the paths between the city buildings. The f + c interface was less efficient than the fixed f + c interface because users had to bend over the workbench to look at occluded areas behind buildings, resulting in slower navigation speed. Among the three interfaces, o + d is the least appropriate interface for this task, since users had to mentally correlate information collected from the two separate windows.

In terms of user preference, fixed f + c and f + c were equally preferred, due to their convenient windows arrangement. However, fixed f + c had some trade-offs in user preference, in the sense that users liked not needing to turn their heads as often as in the f + c, but complained that the context is sometimes cut off from the display. Our observations suggest that users developed strategies to cope with the cutoff situations in the fixed f + c: when users lost the context information, they generally made a fast, extra motion to view the whole context and come back to the original task.

In the map reading task, although we did not find statistical difference in terms of performance, most users preferred o + d, due to the following reasons: (1) the task does not require tight coupling between the two views, (2) the detail window is fixed in a place, providing an up-close view into the detailed scene, and (3) the separate overview window provided an intuitive way to understand current location via the rectangular shape marker and to read labels on the landmarks with a quick glance. On the contrary, the larger context window of the focused-based interfaces containing text labels was spread over a larger physical area of the workbench, which took more time for users to read. Furthermore, under magnification, the focus window blocked the view to the context underneath and thus hampered understanding a user's current location.

It was surprising to see that users zoomed in more in fixed f + c than in o + d, even though their focus and detail windows both provided up-close views to the higher level of detail. We conjecture this is because the fixed f + c interface provided the context area in the periphery, and therefore users were able to use the extended terrain information when searching for map features of the national parks. For example, many of the peaks and mountains were placed along ridges rather than in the middle of a flat area. Therefore, if the task was to identify a mountain, users tended to search along the ridges, using the continuous ridges extended between the focus and the context. In o + d, as there is no context around the detail window, users compromised and avoided zooming in too much, so that they still did not have to look at the overview very often. This hypothetical explanation can be supported by the observation that users turned their head to the overview a lot less frequently during the map reading task than during the path following task as presented in Section 7.4.

The analysis of head motion patterns shows that across the different types of tasks, the o + d interface required users to manage information from the two disjoint views. Meanwhile, in the focus-based interfaces, users were able to keep the focus view in their region of regard no matter whether they were attending the view or not.

8.2 Usability in Terms of Visualization on a 3D Workbench

Throughout the evaluation, we observed that providing a proper view into the 3D information is important in 3D information visualization on a workbench for user performance and for the comfort of use.

Due to the horizontal surface of the workbench, users had to bend over toward the scene, which created several problems in performing both of the tasks. In the path following task, users had difficulty looking into the scene in the farther part of the workbench and the occluded area due to the variable heights of the scene, which in turn affected the performance. In the map reading task, it was sometimes difficult to judge 3D shapes of terrain from the almost vertical viewing direction in all three interfaces. Although users were allowed to move around the workbench and look over the scene from physically different directions, observing a part of a scene that *faces away* from the user was still cumbersome. For example, judging the slope of the north side of a mountain was very difficult. Furthermore, the tethered goggle prevented users from physically walking around the workbench, although they frequently stepped back and forth toward the workbench to attain a proper viewing range. Another problem was neck fatigue after extended time (>20 min) bending over the large workbench. Although the weight of the HMPD (500 g) may have contributed to the discomfort of use, we observed it was mainly due to the users' posture while performing the task. Note here that users did not complain of neck pain during the path following tasks.

To address the above-mentioned problems, we are planning interfaces that will allow the user to change his or her viewing direction of the scene more easily, especially to reach information on the far side of the workbench. For instance, an interface to changing the view direction can be assigned to the nondominant hand, while the dominant hand can be retained for various manipulation (e.g., drawing lines) or visualization (e.g., magnification) actions.

9 Design Guidelines

Based on the evaluation results and the observations, we produced the following design guidelines for 3D information visualization in 3D workbench environments.

1. *f + c interface for multi-layer information visualization:* Consistent with the previous studies (Baudisch et al., 2002); the focus-based interfaces demonstrated superior performance over the o + d interface for a task that requires tight correlation between multiple layers of information. Here, the layers of information can be not only different levels of detail, but also different categories of information, for example, the wiring in a building, a sewage system in a city, or the bone structure of a body relative to organ placement. If a task requires comparably "local" context surrounding the focus area, or if users are generally familiar with the overall structure of a dataset, fixed f + c may be more suitable than f + c. If the task requires studying the whole context area or needs working with multiple focus windows, f + c would be better.
2. *o + d for information-dense visualization, where only a loose level of mental correlation between the information layers is required:* The o + d interface was preferred by many users when the information was densely populated especially with text information and required intensive examination of a local area relative to the very large area. HMDs typically offer limited FOVs relative to the large information space, and thus it is helpful to provide a small overview of the whole information space so that users can build an overall mental model of the

space and keep the location awareness via one quick glance at the overview.

3. *A visualization method to display continuous, coherent transition between focus and context is required even when the two views differ in scale.* During the map reading task, users complained about the degraded correlation between the context and the zoomed-in focus window. Since the typical distortion-based visualization techniques (e.g., fish-eye lens interface) are not practical for 3D visualization, we need to find alternative possible visualization methods to alleviate the discrepancy between the two views caused by the scale differences. One feasible approach may be based on Drag-Mag (Ware and Lewis, 1995), which displays the inset focus window at a slight offset distance and links the focus window with an area in the context using the connected lines. However, this approach is somewhat similar to the o + d, as the inset window is separate from its original location in the context, which may also require some mental correlation similar to the o + d interface. Furthermore, the focus window in the Drag-Mag approach still blocks a part of the context. So f + c interfaces require further study in this direction.
4. *Ability to reach information that is located out of arm's length:* The large extent of workbench displays presents a problem in that it is difficult to view or reach virtual objects placed out of arm's length. Furthermore, the workbench was a physical obstacle for users to walk around, carrying the cable tethered to the helmet. The o + d and fixed f + c in our study provided a detail view centered on the workbench display, which only partially addressed the accessibility problem. Scroll (Smith, 2004) or portal (Khan, Fitzmaurice, Almeida, Burtnyk, and Kurtenbach, 2003) interfaces have been studied for 2D interface in large-scale projection displays to overcome similar problems. In the scroll interface, a user can scroll the display into view via a "wheeling" gesture by one finger. It shares similarity with the fixed f + c interface to some extent in that the context of fixed f + c *scrolls into* the view. Its advantage is that a user does not have to reach for an interface to view the information that is too far from the current location. The "portal" interface is a small inset window showing the part of the scene that is too far out of reach. The menu icons strategically placed around the inset window allow a user to manipulate virtual objects that are placed in a remote location. Such an approach is also used for interaction in an IVE, where the portal windows are used to view or manipulate virtual objects in different coordinate systems (Hirose, Ogawa, Kiyokawa, and Takemura, 2006). For navigation in large-scale virtual environments, Kopper, Ni, Bowman, and Pinho (2006) used virtual landmarks as portals to enter worlds in different levels of scale. Such types of interfaces can be incorporated into a 3D workbench environment, but extra caution should be given to the trade-offs such as the cost of limited workbench display space and increasing interface complexity.
5. *Reorienting information to convenient direction toward a user:* Due to the 3D property of information, there are texts or graphical objects that are oriented "away" from the user's current viewpoint. For example, the slope of a mountainside that faces away from a user is very difficult to see. The text information can sometimes be easily fixed by automatically reorienting it toward a user. However, to do this requires the text information to be stored separately from the graphical objects, and in many cases it is not practical to reenter all the text information for the 3D visualization, as in our map reading task. Furthermore, reorienting a part of the scene is not feasible for graphical objects. An easy solution to this problem may be providing an additional interface device or a widget to allow rotation of the entire scene. However, in general, the addition of devices or widgets complicates the interface, as well as possibly requiring a separate mode change in interaction, which often triggers mistakes by users. Using the paradigm of tangible interface may allow users to intuitively handle the view. For example, Illuminating Clay (Piper, Ratti, and Ishii, 2002) provides a turntable for multiple users to view the terrain from a de-

sired viewpoint, allowing multiple users to naturally collaborate on a shared model.

6. *Supporting various 3D cues:* Stereoscopic viewing is useful in judging 3D depth only to a certain ex tent. Additionally, the horizontal surface of the workbench is problematic in perception of 3D shapes. Taking an analogy from the real world, we may have experienced that it is difficult to tell the exact shape of terrain when we see it from high up in the sky while traveling in an airplane. Diverse 3D visual cues combined with stereoscopy can greatly enhance 3D perception of a scene on workbench displays. For example, the panning behavior in the fixed f + c used in our study presents motion parallax as the scene moves toward a user standing in front of the display.
7. *Workbench design that fits with user's posture and task purposes:* In our evaluations, users complained of neck pain after extended use. It was mainly due to the horizontal orientation of the workbench, which forced users to bend over toward the workbench, especially when users tried to reach information on the far side of the workbench. Such workbench design is common especially in collaborative environments, since it affords an equally advantageous viewing direction to users standing or sitting around the table. Workbench designs with better ergonomic consideration are required. We are planning to redesign our workbench to ergonomically fit with the users' physical posture, but also have it be flexible enough to afford multiple users to comfortably gather around the table.

10 Conclusion

In this article, we presented empirical studies on multi-scale visualization interfaces on a 3D workbench. The three interfaces, f + c, fixed f + c, and o + d, were chosen to test the usability of different window arrangements and navigation methods. Users performed two types of tasks, path following and 3D map reading, which were designed to examine usability issues of the interfaces in a broad range of applications. While the former task mainly evaluated users' ability to mentally integrate the two levels of detail offered by the windows, the latter task considered the readability and accessibility of information spread across the table.

From the evaluation results, users performed the path following task 22% faster in fixed f + c than o + d, and 12% faster in f + c than o + d. Further analysis of the results suggested that fixed f + c provided the most suitable windows arrangement for tasks requiring a tight coupling of multiple layers of information (like the f + c interface), in particular because it also provided an up-close view of the focus scene for faster navigation. Indeed, user preference reflected the performance results, showing that users mostly preferred the focus-based interfaces.

From the 3D map reading task, there was no significant performance difference among the three interfaces. However, differences in scale factor between the focus-based interfaces and the o + d interface suggested that users took advantage of the surrounding context information provided by the focus-based interfaces during searching. In contrast to the path following task, most users preferred the o + d interface for the searching task, mainly because it allowed them to understand the whole area via a quick glance at the small overview, as opposed to the more expansive context view. For all three interfaces, parts of the scene facing away from the user or occluded by other scene objects were both difficult to see and understand in spite of the head tracking capability.

Based on results from the two different types of task, we recommend the use of focus-based interfaces when a tight coupling between the information layers is required, for example, examination of the wiring in a house or the study of human bone structure relative to the placement of organs. Fixed f + c is more efficient than f + c, especially when the context is not too large or users are generally familiar with the whole information area. In the situation where information is densely populated, and if a task does not require a tight integration of different information spaces, then the o + d interface is more suitable.

In the future, we are planning to study multi-scale interfaces for more extended dimensions, for example, volumetric information visualization. That is, we will try to extend the three interfaces for 6 DOF manipulation.

For example, it would be a challenging problem to allow users to easily move a volume of interest in a volumetric overview, while displaying the volume of interest in a separate detail window.

Another line of study is to investigate multi-layer information visualization. Here, rather than concerning a single type of information in different levels of detail, we will examine the use of flexible interfaces to visualize different types of information in different layers. The focus-based interface would provide a well-integrated view into the multiple, coordinated information layers, whereas the o + d interface would be effective for users to build an overall mental model of the information space. Therefore, we need to consider a proper way to concurrently support the multi-window and integrated-information window.

Acknowledgments

This work is partially funded by the National Science Foundation grant awards 05-34777 and 06-44446.

References

Baldonado, M. Q. W., Woodruff, A., & Kuchinsky, A. (2000). Guidelines for using multiple views in information visualization. *Proceedings of AVI 2000,* 110–119.

Baudisch, P., Good, N., Bellotti, V., & Schraedley, P. (2002). Keeping things in context: A comparative evaluation of focus plus context screens, overviews, and zooming. *Proceedings of CHI 2002,* 259–266.

Baudisch, P., & Rosenholtz, R. (2003). Halo: A technique for visualizing off-screen locations. *Proceedings of CHI 2003,* 481–488.

Bederson, B., & Hollan, J. (1994). Pad++: A zooming graphical interface for exploring alternate interface physics. *Proceedings of UIST 1994,* 17–26.

Bier, E., Stone, M., Pier, K., Buxton, W., & DeRose, T. (1993). Toolglass and magic lenses: The see-through interface. *Proceedings of SIGGRAPH 1993,* 73–78.

Bowman, D., & Hodges, L. F. (1997). An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Symposium on Interactive 3D Graphics 1997,* 35–38.

Brown, L., Hua, H., & Gao, C. (2003). A widget framework for augmented interaction in SCAPE. *Proceedings of UIST 2003,* 1–10.

Brown, L., & Hua, H. (2006). Magic lenses for augmented virtual environments. *IEEE Computer Graphics and Applications, 26*(4), 64–73.

Carpendale, M. S. T., & Montagnese, C. (2001). A framework for unifying presentation space. *Symposium on User Interface Software and Technology 2001,* 61–70.

Fisher, R. (1996). Head-mounted projection display system featuring beam splitter and method of making same. US Patent 5572229.

Furnas, G. W. (1986). Generalized fisheye views. *Proceedings of CHI 1986,* 16–23.

Gutwin, C., & Fedak, C. (2004). A comparison of fisheye lenses for interactive layout tasks. *Graphics Interface 2004,* 213–220.

Hirose, K., Ogawa, T., Kiyokawa, K., & Takemura, H. (2006). Interactive reconfiguration techniques of reference frame hierarchy in the multi-viewport interface. *IEEE VR 2006,* 73–80.

Hornbaek, K., & Frokjaer, E. (2001). Reading of electronic documents: The usability of linear, fisheye, and overview + detail interfaces. *Proceedings of CHI 2001,* 293–300.

Hornbaek, K., Bederson, B., & Plaisant, C. (2002). Navigation patterns and usability of zoomable user interfaces with and without an overview. *ACM TOCHI 2002, 9*(4), 362–389.

Hua, H., Gao, C., Biocca, F., & Rolland, J. P. (2001). An ultra-light and compact design and implementation of head-mounted projective displays. *IEEE VR 2001,* 175–182.

Hua, H., Girardot, A., Gao, C., & Rolland, J. P. (2000). Engineering of head-mounted projective displays. *Applied Optics, 39*(22), 3814–3824.

Hua, H., Brown, L., & Gao, C. (2003). A new collaborative infrastructure: SCAPE. *Proceedings of IEEE VR 2003,* 171–179.

Hua, H., Brown, L., & Gao, C. (2004). SCAPE: Supporting stereo-scopic collaboration in augmented and projective environments. *IEEE Computer Graphics and Applications, 24*(1), 66–75.

Khan, A., Fitzmaurice, G., Almeida, D., Burtnyk, N., & Kurtenbach, G. (2003). A remote control interface for large displays. *Proceedings of UIST 2003,* 127–136.

Kijima, R., & Ojika, T. (1997). Transition between virtual environment and workstation environment with projective head mounted display. *Proceedings of VRAIS 1997,* 130–137.

Kopper, R., Ni, T., Bowman, D., & Pinho, M. (2006). Design and evaluation of navigation techniques for multiscale virtual environments. *Proceedings of VR 2006,* 175–182.

LaViola, J., Feliz, D. A., Keefe, D., & Zeleznick, R. (2001). Hands-free multi-scale navigation in virtual environments. *Proceedings of ACM SIGGRAPH 3D 2001.*

Leung, Y. K., & Apperley, M. D. (1994). A review and taxonomy of distortion-oriented presentation techniques. *ACM Transactions on Computer-Human Interaction, 1*(2), 126–160.

Looser, J., Billinghurst, M., & Cockburn, A. (2004). Through the looking glass: The use of lenses as an interface tool for augmented reality interfaces. *Computer Graphics and Interactive Techniques in Australasia & South East Asia 2004,* 204–211.

Melzer, J. E., & Moffitt, K. (1996). *Head mounted displays: Designing for the user*. New York: McGraw-Hill.

Mendez, E., Kalkofen, D., & Schmalstieg, D. (2006). Interactive context-driven visualisation tools for augmented reality. *Proceedings of ISMAR 2006,* 209–216.

North, C., & Shneiderman, B. (2000). Snap-together visualization: Evaluating coordination usage and construction. *International Journal of Human-Computer Studies* (Special Issue on Empirical Studies of Information Visualization), 53(5), 715–739.

Oh, J.-Y., & Hua, H. (2006). User evaluations on form factors of tangible magic lenses. *Proceedings of ISMAR 2006.*

Pierce, J. S., & Pausch, R. (2004). Navigation with place representations and visible landmarks. *Proceedings of IEEE VR 2004,* 173–288.

Piper, B., Ratti, C., & Ishii, H. (2002). Illuminating clay: A 3-D tangible interface for landscape analysis. *Proceedings of CHI 2002,* 355–362.

Plaisant, C., Carr, D., & Shneiderman, B. (1995). Image-browser taxonomy and guidelines for designers. *IEEE Software, 12*(2), 21–32.

Reilly, D., Rodgers, M., Argue, R., Nunes, M., & Inkpen, K. (2006). Marked-up maps: Combining paper maps and electronic information resources. *Journal of Personal and Ubiquitous Computing, 10*(4), 215–226.

Schmalstieg, D., & Schaufler, G. (1999). Sewing worlds together with SEAMS: A mechanism to construct complex virtual environments. *Presence: Teleoperators and Virtual Environments, 8*(4), 449–461.

Smith, G. M. (2004). The radial scroll tool: Scrolling support for stylus- or touch-based document navigation. *Proceedings of UIST 2004,* 53–56.

Stoakley, R., Conway, M. J., & Pausch. R. (1995). Virtual reality on a WIM: Interactive worlds in miniature. *Proceedings of CHI 1995,* 265–272.

Ullmer, B., & Ishii, H. (1997). The metaDESK: Models and prototypes for tangible user interfaces. *Proceedings of UIST 1997,* 223–232.

Viega, J., Conway, M. J., Williams, G., & Pausch, R. (1996). 3D magic lenses. *UIST 1996*, 51–58.

Ware, C., & Lewis, M. (1995). The DragMag image magnifier. *Proceedings of CHI 1995,* 407–408.

Wingrave, C. A., Haciahmetoglu, Y., & Bowman, D. A. Overcoming world in miniature limitations by a scaled and scrolling WIM. *Proceedings of 3DUI 2006,* 11–16.

Appendix A. User Attention Categorization

Based on the literature on eye and head motion ranges, in general, the comfortable eye movement range is about ±12° from the center with the head fixed in one position (Melzer and Moffit, 1996). To attend an area of interest more than 12° away, people tend to turn their head toward the interest area rather than turning their eyes. In addition, the foveal vision of the eye, which provides high resolution for resolving detailed information, is only about 2–3°. Based on those previous study results, we set up two types of thresholds. The first threshold is set to be 15°, which is the added sum of the eye's foveal vision (2–3°) and the comfortable eye motion range (12°) when the head is fixed. We use this threshold to decide if a display window is completely *out* of a user's current attention. If the nearest border of a display window is more than 15° away from the head's center of focus, we conjecture that the window is completely out of attention. The second threshold is set to be 5°, which is used to decide if a window is mainly attended by the user. When the head's center of focus is *inside* the area of a window and is more than 5° away from the nearest border of the window, we conjecture that the user's main focus is inside the window itself rather than observing outside of it. Along with the inside and outside conditions, there are always points where a head's center of focus cannot belong to either condition, which is inbetween the inside and outside of attention. That is, if we observe that the head's center of focus is on the border between micro and macro interface windows, we conjecture that the user is trying to observe different levels of information from the two interface windows.

We calculated the intersection point between the workbench surface and a ray subtended from the user's head position along the head's orientation in 3D space. The intersection point was designated as the center position of the user's attention on the workbench, which is hereafter referred to as the user's gaze center for simplicity. Based on the two above-mentioned thresholds, in the two focus-based interfaces, we decide a user is looking into (*IN*) the focus window if the user's gaze center is inside the window more than 5° away from all the borders of the focus window. If the gaze center is more than 15° away from all the borders of the focus window, we decide the user is looking away (*OUT*) from it. All positions in between are categorized as *BOTH,* meaning the user is trying to look at both of them to integrate the two views. In the o + d interface, using the *IN* threshold (5°), if the user's gaze center position is inside the detail window, it is categorized as *IN*. If the gaze center is inside the overview window, it is categorized as *OUT*. All positions in between are categorized as *BOTH*.

Appendix B. Sample Questions for 3D Map Reading Task

A. 1 Glacier National Park

One-area questions:

From the area between Many Glacier Information and Saint Mary Lake, find mountains: Mt Gould, Going-to-the-Sun Mtn. Which one is located where a ridge ends?

In the area south of Saint Mary Lake and east of Lewis Range, find these creeks: Medicine Owl Creek, Red Eagle Creek. Which one is steeper?

Two-area questions:

Find these peaks/mountains: From northwest of Lake McDonald, Longfellow Peak; from east of Lewis Range and south of Saint Mary Lake, Divide Mountain. Which one is located where multiple ridges meet?

Find these creeks/rivers: Run from north side of Saint Mary Lake, Rose Creek. In the northeast of Livingston Range, Waterton River. Which one is steeper?

A. 2 Olympic National Park

One-area questions:

In the area east of Elwha River, find peaks/mountains: Obstruction Peak, Green Mt. Which of them is located where multiple ridges meet together?

From the east side of Elwha River, find creeks: Morse Creek, Maiden Creek. Which one is steeper?

Two-area questions:

Find places: East of Hurricane Ridge Visitor Center, Steeple Rock; north of Sol Duc River, Sourdough Mountain. Which one is steeper on its west side?

Find creeks: East of Livingston Range, Mineral Creek, south of Lake McDonald, Snyder Creek. Which one is steeper?