Navigation and Selection in 3D User Interfaces

Matthew Ruby
User ID: dcs2mr
Student ID: 000383117
E-mail: matthew.ruby@durham.ac.uk
March 19th 2003

Abstract

Improvements in computer hardware and software have led to great advancements in 3D graphics and interest in their application in the development of graphical user interfaces (GUI). While various techniques are employed to create the illusion of 3D it is inherently difficult to effectively communicate spatial relationships within the 2D medium of a flat computer screen. The designers of 3D interfaces must consider the cognitive aspects of navigation and selection that involve the development of mental maps from components of the environment through the interpretation of affordances and spatial information. The illusion of 3D depends on the perception of depth cues within the display space that provide a structure from which a cognitive map can be formed. Navigation metaphors determine how interaction occurs within a virtual environment according to a user's spatial ability and must, therefore, be appropriate for the user and the tasks that he seeks to accomplish.

The goal of interface design is to reduce the cost structure of a virtual environment to allow a user to complete a task with minimal effort. In a 3D display, problems such as occlusion, the distortion of perspective, visual complexity and user disorientation increase the cost structure and need to be minimized through the use of appropriate presentation techniques and user interaction. Focus+context techniques and selective dynamic manipulation may be employed to achieve this in a 3D display space.

Table of Contents

1 Introduction

Advancements in computer hardware and software have led to great interest in the development of 3D graphical user interfaces (GUI). While 3D GUI's have the potential to allow users to interact with computers in new and exciting ways the great challenge for interface designers is to realistically depict objects in 3D space on the 2D medium of a flat computer screen. It is inherently difficult to effectively communicate spatial relationships within a 2D medium so that a user is able to navigate and interact with the environment.

Cognitive aspects of human perception must be considered in the design of any computer interface. The development of 3D displays requires particular attention to the impact of human spatial ability and the perception of depth cues on navigation within this environment. Navigation depends on the development of wayfinding strategies; a process that involves building mental models based largely on spatial cues and affordances provided by the display space.

Designers of 3D interfaces face certain issues that are unique to this environment including the effects of occlusion and visual perspective, visual complexity, and user disorientation that are a direct result of the additional dimension and can be a detriment to effective navigation and interaction. Techniques that reduce the cost structure of the display space must be considered to address these issues.

2 Cognitive Issues

This section considers the cognitive aspects of human perception and their influence on interface design. Affordances and spatial cues form the cognitive basis for the illusion of 3D and the formation of mental maps that allow a user to develop effective navigation strategies and interact with a 3D environment.

2.1 Navigation

Navigation is the planning and execution of travel through space. To navigate an environment, whether a GUI or a physical space, a user must establish a means of reference to the environment so that he remains oriented and does not become lost. As we have a relatively limited ability to utilize physical cues within the environment the human perceptual system has evolved various means to facilitate navigation through wayfinding strategies that involve the formation of cognitive maps [DARK93].

2.1.1 Wayfinding and Cognitive Maps

Ware [WARE00] defines wayfinding as the process of building up an understanding of one's environment through the formation of cognitive mental models and through the use of physical maps. Darken and Sibert [DARK93] suggest that cognitive maps are formed from basic components including paths, edges and landmarks that correspond to objects within the environment and nodes and districts that represent logical groupings of these components. Research has shown that cognitive maps are most easily developed when a user has an overview of the entire environment that allows them to establish relationships between components [WARE00]. Supporting wayfinding in a 3D environment presents a challenge for interface designers when it is possible to view objects from multiple perspectives as it can be difficult to create landmarks that are recognizable from all possible points of view.

Physical maps aid navigation when they allow a user to establish a relationship between the environment that they represent and their cognitive map. They can actually help to form a cognitive map more quickly than would be possible by exploring an environment, especially when matching points on an overview and detailed view are shown to overcome perceived scaling errors. Cognitive load is reduced when an overview map of a large information space is provided showing user location and orientation as well as key landmarks.

2.1.2 Spatial Ability

A user's spatial ability can be a significant determinant of how well they are able to use a computer system [BENH99]. This is especially true in 3D environments where the additional dimension of depth places greater demands on spatial ability to navigate and interact with the interface. Satalich [SATA95] categorizes spatial ability as spatial orientation, spatial visualization and spatial relations. Spatial orientation is the ability to retain relationships between objects and use oneself as a reference while manipulating them. Spatial visualization involves the ability to manipulate the relationships within an object, and spatial relations is concerned with the ability to create a mental image of an object from different imagined view points. Each of these components must be considered in relation to the spatial ability of a system's users when designing a 3D interface.

2.2 Affordances

According to affordance theory, *affordances* are properties of the environment that we perceive for the purpose of action [WARE00]. In terms of interface design, affordances are the ways that interface components communicate how they may be used to allow a user to accomplish a task. Affordances are, to a great extent, borrowed from the real world in that

they are generally pictorial representations of every-day objects so their perception is significantly influenced by a user's experience and convention [SPEN00]. The goal for an interface designer is to ensure that perceived and actual affordances are identical and to provide affordances that are intuitive so that users are able to easily grasp their meaning and understand their utility.

Affordances for 3D interface components have often been borrowed from 2D interface design but they should ideally reflect the dimensionality of the 3D environment and suggest interaction that takes advantage of this. Affordances that are developed with this in mind should be intuitive for most users as we are very accustomed to three dimensional manipulation and navigation in the real world.

2.3 Depth Cue Theory

Depth cue theory is the study of how information about the 3D world is provided via various depth cues that are translated by the human perceptual system to obtain an understanding of physical space. This is essential to realistic portrayal using 3D graphics where the same information must be provided visually to maintain the illusion of three dimensions. A user's ability to perceive the actual size of an object is a measure of the effectiveness of depth cues in a 3D interface [WARE00]. This section describes some of the most effective and most commonly used depth cues in 3D interface design.

2.3.1 Perspective

The geometry of perspective defines the extension of rays from a viewpoint to features in the environment. Depth cues are provided by the convergence of parallel lines to a single point, by the relative difference in size of distant objects compared to those that are close to the viewer (perspective distortion) and by the perception of scale afforded by the proximity of objects to other objects whose size is known (Figure 2.1). The elements of uniformly textured surfaces can also provide effective depth cues as they will become smaller as they become more distant from a viewpoint [WARE00]. The use of perspective in 3D interfaces supports the perception of distance as well as direction and can be a very effective way to ensure that a user remains oriented within the virtual environment.

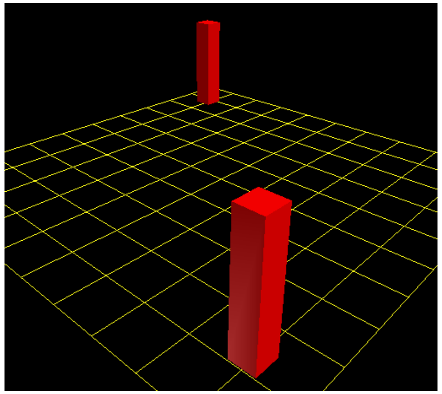

Figure 2.1 Perspective distortion and depth cues from parallel lines and surface texture.

2.3.2 Occlusion

Occlusion, where one object overlaps another, provides a very strong depth cue in that the viewer perceives one object to be closer to them in relation to the other. Occlusion does not in itself provide any information concerning the distance between objects or their relative size but can be very effective depth cue for discerning the structure of visually complex objects. Transparency is a means of partial occlusion that can be used to provide a depth cue while allowing information about the occluded object to be revealed. Rekimoto and Green's Information Cube (Figure 2.2) [REIK93] uses 3D semi-transparent cubes to display hierarchical data as a series of nested boxes. Objects are shown in context with other items in the information space but there is a trade-off between increased visibility and visual complexity. Reikimoto and Green solve this problem by increasing the opacity of cubes that are deeper within the hierarchy so that detail is only revealed as a user moves closer to his intended target.

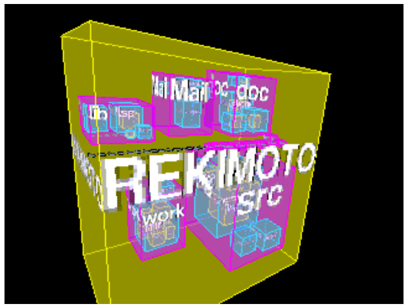

Figure 2.2 Transparency as a form of partial occlusion in Rekimoto and Green's Information Cube.

2.3.3 Cast Shadows

Cast shadows are caused by the absence of light created by an occluding object between a light source and a surface. They can provide very effective depth cues that help to establish the height of an object above a plane and locate objects in relation to a surface. Studies have shown that cast shadows affect the perceived size, elevation and relative depth of an object and can provide additional information about shape, layout and depth [HUBO02]. The effect is most pronounced when shadows are cast on a surface that is relatively close to the user as a user is more likely to be able to connect an object with its shadow.

Robertson et al [ROBE91] use cast shadows in their Cone Tree implementation (Fig 2.3) to provide additional cues regarding the overall structure of the tree as it is manipulated within the information space.

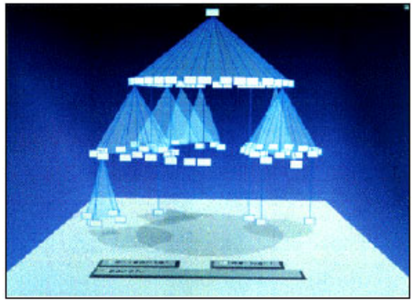

Figure 2.3 Cast shadows in Robertson et al.'s Cone Tree.

2.3.4 Shape from Shading

Depth cues from shading, as with cast shadows, originate from the interaction of light with surfaces. Shading techniques are categorized into four basic models:

Lambertian shading:
Light reflected from an object equally in all directions based on surface geometry. The brightest surfaces are those that face the light source.

Specular shading:
Highlights reflected from a glossy surface. Shading depends on the shininess of the surface and the proximity of the view point to the point of reflection.

Ambient shading:
Environmental light that arrives equally from all directions.

Shading is used in a 3D interface to reveal surface shape that defines the structure of objects. Selection of the appropriate shading models is application specific and depends on the combination of shading that best simulates structural depth cues.

2.3.5 Structure from Motion

As objects move through space they provide varied feedback to a viewer that is perceived as patterns of light. *Motion parallax* occurs when a viewer travels past objects at varying distances, akin to looking sideways out of a moving vehicle. Distant objects appear to move more slowly than those that are closer to the viewer. When a viewer moves forward through a cluttered environment an expanding pattern of motion known as a *velocity field* is perceived. *Kinetic depth effects* occur when objects move through 3D space and thereby present different profiles to the viewer. The brain assumes that objects are rigid so movement reveals structural information that provides insight as to their shape (Figure 2.4).

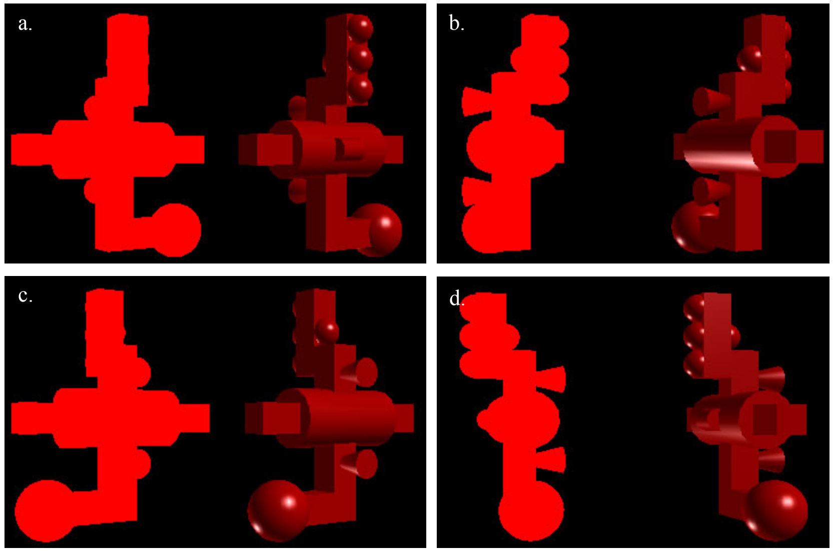

Figure 2.4 Rotation sequence showing structure from motion for the object at left.

2.3.6 Stereoscopic Depth

Stereoscopic depth cues originate from the difference in viewpoint relative to the distance between the eyes. Each eye views a slightly different image and the brain interprets differences in the placement of objects by assuming that they are at different depths. This is illustrated by Figures 2.5 and 2.6 where objects that fall on an imaginary arc know as the

horopter are perceived to be at equal distances from the viewer. The angle formed by lines drawn from an object to each retina will be equal for objects that are at equal distance from the viewer and more acute or obtuse for objects respectively further from and closer to the viewer.

Stereoscopic displays create the illusion of depth using these principles by dividing the display into sections to present slightly offset views of an image to each eye. The two different views are then presented on the screen, one to each eye, and the images appears in stereo. This creates a very realistic illusion of depth that can greatly enhance a user's sense of presence within a 3D environment.

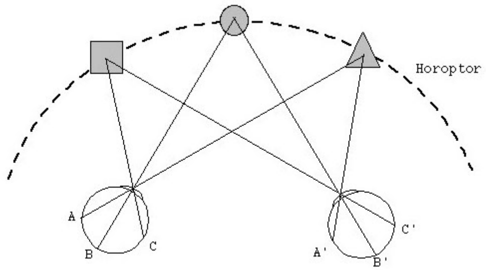

Figure 2.5 Perception of objects at equal distances from the viewer.
Source: www.cs.dartmouth.edu/~wittie/reu/stereo.html

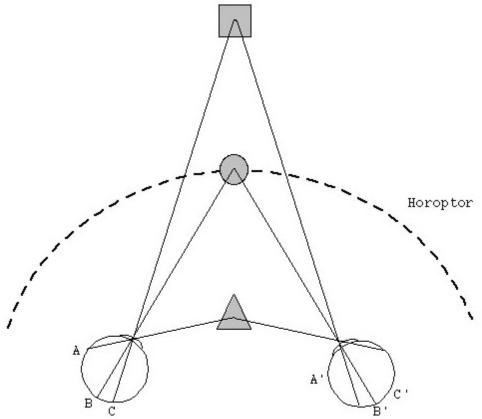

Figure 2.6 Perception of objects at different distances from the viewer.
Source: www.cs.dartmouth.edu/~wittie/reu/stereo.html

3 Interface Design

In addition to providing appropriate affordances and environmental cues, 3D interfaces must consider appropriate methods for navigation and user interaction that specifically address issues associated with the additional dimension. Various techniques have been developed to mitigate these problems and help to reduce the cost structure of the display space in 3D environments. These include metaphors that facilitate effective navigation, focus+context presentation methods that help to maintain orientation, and manipulation and selection techniques that address occlusion and the distortion of perspective.

3.1 Navigation Metaphors

Navigation metaphors are cognitive models for interaction with an environment that determine how a user interacts with the display space by adjusting the viewing position. The metaphor chosen will condition a user to expect certain behaviour from the environment and determines how easy or difficult it is to accomplish certain tasks [WARE00]. It is therefore essential that the navigation metaphor is appropriate to the work that a user seeks to accomplish.

In the *world in hand* metaphor a user grabs a part of the environment to manipulate it. To zoom in a user pulls part of environment towards him, to zoom out user pushes part of the environment away from him and to rotate the environment a user twists it. This metaphor is best for viewing relatively compact environments as there are few cues for navigation over long distances.

In the *eyeball in hand* metaphor the user directly manipulates their viewpoint to observe the environment from a different perspective. This is usually not particularly effective as certain viewpoints may be confusing or impossible to achieve when constrained by the boundaries of the display or by objects that block their path.

A *walking* metaphor is often used in virtual reality applications or in information landscapes where a user is able to move through the environment constrained only by the effects of virtual gravity that maintains the viewpoint at a certain distance above a plane. Although the effect of gravity does restrict movement it can help to prevent disorientation as the user is always able to determine the up direction.

The *flying* metaphor allows a user to move through an environment as if they were in an aircraft but without the restrictions of gravity. This is a very flexible navigation metaphor as there are usually minimal constraints in any direction and a wide range of movement and perspectives are achievable.

3.2 Cost of Knowledge

The idea that information systems have a cost structure was proposed by Card et al. in [CARD94] where they define a Cost-of-Knowledge Characteristic Function. This is a calculation of the additional useful information that becomes available as a user navigates an information structure for each unit of time expended. The more time that a user spends navigating through successive information structures within a system the higher the cost in relation to information discovered. A similar cost/benefit analysis was proposed by Furnas in [FURN97] where he explores the requirements for what he calls the efficient view traversibility of very large information structures. The goal is efficient navigation of information structures within the constraints of limited screen space and time. He characterizes the visualization of an information structure as a view graph with nodes representing components of the information structure logically connected by all possible navigation paths between them. Efficient view traversibility is a measure of how navigable a view graph is: it is efficiently view traversable if the user has a manageable number of choices at any given node and the distance between nodes is small in relation to the size of the overall structure. Furnas also introduces the concept of view navigability which is a measure of how readable the view graph is in terms of the user's ability to find the shortest path to their goal.

3.3 Navigation and Selection Techniques

The cost of knowledge is a concept that can be used as a guideline for improving interface design by developing environments that minimize the amount of time and work a user must invest in performing a task. Various techniques have been developed for use in 3D interfaces to accomplish this that allow a user to efficiently traverse the information space and remain oriented within the environment.

3.3.1 Focus+Context

Focus+Context techniques provide both overview and detail by integrating both in the same view using distortion or selective reduction techniques that allow a user to focus on specific components. The idea was first proposed by Furnas [FURN81] where he observes that when faced with large amounts of information people have a tendency to focus on areas that are of immediate interest and ignore details in the periphery, akin to a fisheye lens. Furnas suggests that it is possible to determine how important a component of the information structure is to the viewer by calculating the *degree of interest*; a measure that combines the significance of the component globally (or level of detail) and its proximity to the current area of interest. In addition to providing focus within visually complex environments focus+context techniques can help a user to remain oriented as he has reference to the entire display space. This is particularly important in 3D worlds where a user has more space in which to become lost.

Distortion techniques have often been analogized as a rubber sheet populated with information that is stretched on a frame [LEUN94]. Stretching the sheet causes some areas to expand and become more detailed while others shrink and become less detailed according to a mathematical transformation function. This presents some unique challenges in a 3D environment where it often becomes necessary to access the internal structure of an object due to occlusion by other components within the display space. Stripping away the occluding objects allows a user to focus on the objects revealed but causes a loss of context with the components removed. One approach to distortion is to change the size of objects within the display area according to the degree of interest calculation. This alone may not have the desired effect in a 3D environment as occluding objects may cause the area of focus to be hidden if it is deep within an information structure. Sheelagh et al. [SHEE97] use the degree of interest calculation in combination with a visual access distortion function to remove occlusions by clearing a line of sight to a focal point within a 3D structure (Figure 3.1).

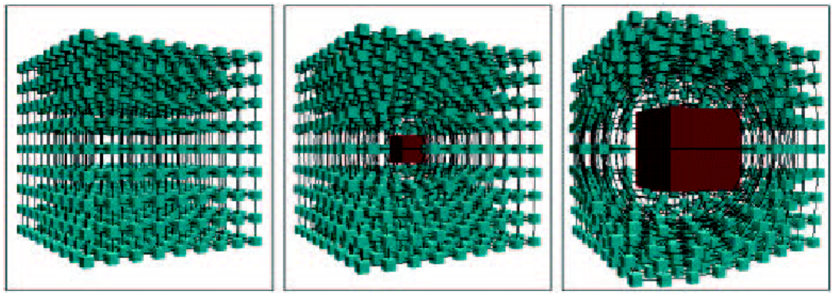

Figure 3.1 An illustration of Sheelagh et al.'s visual access distortion function.

Perspective can be used in a 3D interface to provide a more intuitive view of a virtual environment than other distortion techniques. Card et al.'s [CARD94] Perspective Wall (figure 3.2) uses 3D perspective to allow a user to focus on specific information while maintaining context with the entire information space. Converging parallel lines provide depth cues that suppress the perception of any distortion by creating the illusion of distance.

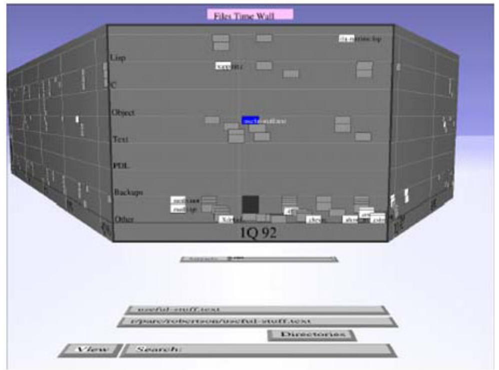

Figure 3.2 Perspective used for focus+context in Card et al.'s Perspective Wall.

In addition to distortion, filtering and selective aggregation are focus+context techniques that may be used to selectively reduce information outside of the focal area. Filtering removes elements of the display using a degree of interest calculation or dynamic queries while selective aggregation causes similar structures to coalesce when they are distant from the area of focus and to dissolve into their component cases when are closer to it.

3.3.2 Manipulation and Selection

The ability to effectively manipulate elements is often necessary to overcome problems of occlusion and the distortion of perspective in 3D environments. Rotation is a manipulation technique that allows the user to effectively adjust the viewing angle and discover new information about the content and structure of an object or group of elements in a display space. Cone Tree implementations use interactive animation to rotate tree structures at a speed that allows a user to track the transition [ROBE91]. This helps to reduce cognitive load by allowing the perceptual system to follow the changes in spatial relationships caused by rotation ensuring that the user remains oriented within the hierarchy.

Precise quantitative comparison of objects in a 3D interface that are shown at different distances from the viewpoint is not feasible without some sort of manipulation due to scaling effects related to perspective. Chuah et al.'s [CHUA95] Selective Dynamic Manipulation system (SDM) demonstrates various techniques to accomplish this and overcome occlusion issues including the ability to group elements into a comparison set. The selection and manipulation of individual elements is accomplished by using object handles that have various functions including the addition of an object to a group, highlighting an object by changing its size or colour or moving an object to a different location (Figure 3.3). Objects that are added to comparison sets can be moved as a group to different locations in the display space to overcome perspective distortion when comparing objects and to overcome occlusion problems (Figure 3.4). Ghost images are displayed in the original location for objects that are moved to preserve spatial relationships and ensure context is maintained.

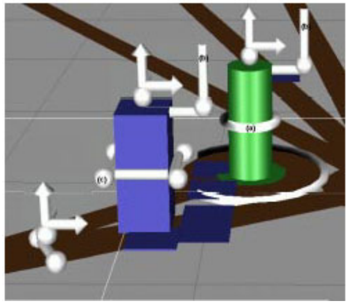

Figure 3.3 Object handles in Chuah et al.'s SDM.

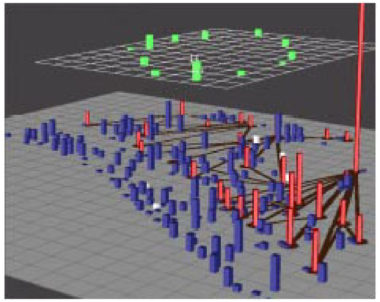

Figure 3.4 Moving a comparison set in Chuah et al.'s SDM.

3.4 3D Versus 2D

The question of whether 3D displays are an improvement over their traditional 2D counterparts is a topic of active research and great debate. Most studies have shown that there is no significant difference in the ability of a user to perform a task in either type of interface but users often report a marked preference for 3D displays [COCB01].

In comparison to 2D, 3D displays do have certain advantages for visualization including a more efficient use of space, especially when effective focus+context techniques are used. The third dimension allows information to be displayed along an additional axis that effectively extends the display area. Cone Tree implementations illustrate this point as they are able to present many more nodes than traditional 2D tree structures within a similar display space. The additional dimension can also allow a user to move to a different view point to gain a better perspective and a deeper understanding of the information displayed including structural relationships and patterns. 3D displays can also present a more realistic simulation of space than their 2D counterparts with depth cues based on real-world metaphors. Whether this provides a more intuitive environment based on human spatial ability remains a subject of research and has not been conclusively determined [COCB01]. Perhaps the greatest potential achievement for 3D interfaces is their ability to afford of a sense of presence within the virtual environment where a user has the sense of being in the location portrayed by the

display. This connection with the environment allows a user to perform efficient navigation and manipulation within the space.

As discussed in previous sections, occlusion and scaling issues related to 3D perspective are problems that are inherent in 3D environments. These problems can be alleviated to a great extent by effective manipulation mechanisms but this places additional demands on a user who must learn to interact with and control the environment in an additional dimension. A major criticism of 3D displays is that they are not truly 3D in that they rely on 2D input and output devices for communication. Within a 3D environment this can cause navigational conflicts as a user may perceive conflicting depth cue information [HERM00]. While 3D displays can be designed to present a more realistic display of space than their 2D counterparts, users are generally much more familiar with 2D interfaces and may require more time to become adept at navigating and manipulating a 3D environment. A user is also more likely to become disoriented in a 3D interface as it is much easier to become lost in three dimensions than it is in two [BENH99]. This can be alleviated to some extent by using focus+contect techniques that allow a user to maintain visual contact with structures that serve as landmarks.

4 Conclusion

Navigation of any space requires a mental model of various features within the environment. A user's ability to navigate within a 3D interface therefore depends on the perception of affordances and spatial cues that provide the information needed to develop a cognitive map. Affordances must be developed to convey meaningful information and spatial cues provided that are based on our understanding of human depth perception and spatial ability. Navigation metaphors determine how interaction occurs within the display space and must therefore be appropriate for the tasks that a user seeks to accomplish within the environment.

The goal of interface design is to reduce the cost structure of a virtual environment to allow a user to complete a task with minimal effort. In a 3D display this depends on methods that mitigate problems such as occlusion, the distortion of perspective, visual complexity and user disorientation. Effective focus+context techniques can help a user remain oriented by providing an overview of the display in which landmarks remain visible and reduce visual complexity by creating areas of focus, through selective filtering or by selective aggregation. Manipulation techniques can allow a user to move objects in the display to deal with occlusion and to highlight or relocate objects for the purpose of comparison to overcome the distortion caused by perspective.

There is no definitive answer to the question of whether 3D interfaces are an improvement over their 2D counterparts. While 3D graphics can be used to develop very effective interfaces the appropriate technique to use will depend on what method is best suited to the task that a user seeks to accomplish.

5 References

[BENH99] BenHajji, F. and Dybner, E. 3D Graphical User Interfaces. Department of Computer and Systems Sciences Stockholm University and The Royal Institute of Technology July 1999.

[CARD94] Card, S., Pirolli, P. and Mackinlay, J. The Cost-of-Knowledge Characteristic Function; Display Evaluation for Direct Walk Information Visualizations. Proceedings of CHI'94, 1994, pp. 238-244.

[CARD99] Card, S., Mackinlay, B., Shneiderman, B. Readings in Information Visualization: Using Vision to Think. Morgan Kaufman, 1999.

[CHUA95] Chuah, M., Roth, S., Mattis, J. and Kolojejchick, J. SDM: Selective Dynamic Manipulation of Visualizations. Proceedings of UIST '95, Pittsburgh, PA, pp. 61-70, 1995.

[COCB01] Cockburn, A. and McKenzie, B. 3D or not 3D?: Evaluating the Effect of the Third Dimension in a Document Management System. SIGCHI : ACM Special Interest Group on Computer-Human Interaction. Pp. 434-441, 2001.

[DARK93] Darken, R.P., & Sibert, J.L. A Toolset for Navigation in Virtual Environments. Proceedings of the ACM User Interface Software and Technology, pp.157-165, 1993.

[FURN81] Furnas, G. The Fisheye View: A New Look at Structured Files. AT&T BEl Labs Technical Report, 1981.

[FURN97] Furnas, G.W., Effective View Navigation. Proceedings of CHI '97, Atlanta, GA, April 1997, pp. 367-374.

[HERM00] Herman, I., Melancon, G. and Marshall. M. Graph Visualization and Navigation in Information Visualization: A survey. IEEE Transactions on Visualization and Computer Graphics, 6(1) pp. 24-43, 2000.

[HUBO02] Hubona, G.S., Shirah, G.W. and Jennings, D. The Use of 2D and 3D Depth Cues in Performing Computer-Generated Spatial Tasks. www.cis.gsu.edu/~ghubona/ieeesmc/hubona1.pdf

[LEUN94] Leung, Y. and Apperley, M. A review and taxonomy of distortion-oriented presentation techniques. ACM Transactions on Computer-Human Interaction, Vol. 1, No. 2, June 1994, pp. 126-160.

[REIK93] Rekimoto, J. and Green, M. The Information Cube: Using Transparency in 3D Information Visualization. Proc. of Third Annual Workshop on Information Technologies & Systems (WITS'93) (December 5, Orlando, Florida, USA), pp. 125-132.

[ROBE91] Robertson, G., Mackinlay, J and Card, S. Cone Trees: Animated 3D Visualizations of Hierarchical Information. Proc. CHI '91 : Human Factors in Computing Systems, pp. 189-194, ACM Press.

[ROBE93] Robertson, G., Card, S. and Mackinlay, J. Information Visualization Using 3D Interactive Animation. Communications of the ACM, Vol. 36, No. 4, Apr. 1993, pp. 57-71.

[SATA95] Satalich, G.A. Navigation and Wayfinding in Virtual Reality: Finding the Proper Tools and Cues to Enhance Navigational Awareness. MS Thesis, www.hitl.washington.edu/publications/satalich, 1995.

[SHEE97] Sheelagh, M., Carpendale, T., Cowperthwaite, D. and Fracchia, F.D. Extending Distortion Viewing from 2D to 3D (1997) IEEE Computer Graphics and Applications: Special Issue on Information Visualization.

[SPEN00] Spence, R. Information Visualization. ACM Press, 2000.

[WARE00] Ware, C. Information Visualization. Morgan Kaufmann, 2000.