The World Processor
An Interface for Textual Display and Manipulation in Virtual Reality

by

Jouke C. Verlinden, Jay D. Bolter, and Charles van der Mast

GIT-GVU-93-13
April 1993

Graphics, Visualization & Usability Center

Georgia Institute of Technology
Atlanta GA 30332-0280
The World Processor
An Interface for Textual Display and Manipulation in Virtual Reality

Jouke C. Verlinden  Jay David Bolter  Charles van der Mast
GVU Center  GVU Center  Department of Information Systems
Georgia Institute of Technology  Georgia Institute of Technology  Delft University of Technology

ABSTRACT
We propose a Virtual Reality interface for creating and manipulating textual and hypertextual information. Many current hypertext systems include concept maps in two dimensions. Our system, called the World Processor, offers the user a three-dimensional concept map. It allows the user to move through the map while exploring existing objects and their associated texts and adding new objects and texts. We discuss a number of interface issues raised by three-dimensional hypertext and then describe the World Processor as an attempt to address some of these issues. Of particular interest in the World Processor interface are its point-and-shoot selection technique, its methods for reading and editing text in 3D, its widgets for creating and modifying conceptual objects, and its auditory as well as visual feedback. Possible applications for three-dimensional hypertext include CSCW (particularly collaborative writing) and authoring systems for computer-based training.

KEYWORDS: Hypertext, Virtual Reality, information visualization, interactive graphics.

INTRODUCTION
Current Virtual Reality systems typically offer their users simple, perceptual environments. These environments do not include abstract, symbolic information; often they are entirely devoid of textual messages or even labels. Such environments are perfectly well suited to many applications in simulation and telepresence. Yet there is no theoretical reason why Virtual Reality (or three-dimensional graphics in general) should be limited to nontextual applications. Human beings have well-developed skills in navigating and interacting in three-dimensional environments, and these skills can be exploited to give significance to abstract information when it is spatially displayed. Work in scientific visualization and information visualization shows how abstract or symbolic information can be conveyed in three dimensional spaces. Our goal is to extend these existing techniques; we propose an interface specifically designed to display and manipulate textual and hypertextual information in Virtual Reality.

SPATIALIZING TEXTUAL INFORMATION
The idea of putting textual information in a three-dimensional space may seem radical. We usually think of writing as restricted to various special surfaces (e.g. paper, whiteboard, or computer screen). However, there is a long historical tradition of labeling and writing on objects in the world. Clay vessels, stone tablets, the walls and doors of buildings, metal road signs, and so on have all served as places for writing [7]. There is also a long tradition of locating semantic objects in an imagined space. Orators in ancient Greece and Rome used spatialization to memorize speeches by a technique called "the art of memory": they remembered each speech as an imaginary walk through a three-dimensional structure, usually a building or a garden [26].

Computer graphics, especially three-dimensional graphics and Virtual Reality, provide contemporary technology that can vastly extend the art of memory and help users imagine semantic objects in space. Scientific visualization has already become a recognized discipline for semantic representation in space. In applications for scientific visualization, the semantic objects represent numerical data. Simple iconic semantics (such as variations in the size, color, and shape of objects called "glyphs") show relationships among complex, multi-dimensional data sets. Some applications allow the scientist to interact with the data -- by navigating to change point of view and by manipulating certain filters. Important work is also being done in
visualizing computer algorithms: here the semantic objects stand for data structures and steps in the algorithm [22].

Spatial representation has also been applied to the representation of verbal ideas and text itself. A pioneering application is this area was the SemNet system in 1986 [6]. SemNet displayed semantic networks in three dimensions. The semantic network appeared as clusters of cubes and lines that were deployed in a virtual space. The cubes represented nodes of information (each cube had a title floating in front of it); the lines established semantic relationships among cubes. The environment was presented to the user on a conventional CRT display, and the user could interact with the data using a keyboard and mouse. The Xerox PARC project called 3D/Rooms has extended this concept of spatialization to "information visualization" in general [5]. Several new visualization tools and interface techniques have been introduced in this project. So-called "cone trees" display hierarchical information (e.g. a directory structure) [21]; temporal or other linear information can be visualized with the "perspective wall," a three-dimensional timeline [13]. In addition to the work at PARC, there is the 3D-based user interface described by Hemmje at the Integrated Publication and Information Systems Institute (IPS) in Darmstadt [9]. Such research suggests that three-dimensional representation allows the information space to be both larger and denser. More information can be conveyed, because the human ability to perceive spatial structures reduces the cognitive load. Perception and action are tightly coupled [8]; animation and manipulation of the structure may also augment the user's understanding of the information represented [5].

INTERACTING WITH HYPERTEXT IN 3D

Hypertext is an information structure in which the user often experiences cognitive overload—the problem of being lost in hyperspace [18]. So three-dimensional representation should improve the display and the manipulation of hypertext. A hypertext can be understood as a collection of nodes connected in a two-, three-, or n-dimensional structure, and some hypertext applications already make this structure visible by allowing the user to build concept maps. In these maps the nodes are often represented by labelled boxes and the links by arrows (Figure 1). One such application is Storyspace [1]; other systems that provide concept maps include MacWeb [17] and the SEPIA system [23]. However, most current hypertext applications provide only two-dimensional concept maps.

In three dimensions, the PARC 3D/Rooms system and the system described by Hemmje offer maps of hierarchical structures and point the way to hypertext. The SemNet system is in fact hypertextual, in the sense that a semantic net is a hypertext. All of these systems focus on the display and exploration of large information spaces. They would be classed as exploratory hypertexts, while in constructive hypertext, the user or reader can modify both structure and text [12].

Figure 1: A Storyspace concept map

Our interface is designed to permit the user to modify text and structure, rather than merely explore them. This interface could be employed wherever constructive hypertext is called for. Three-dimensional hypertext would be appropriate for some kinds of CSCW, for example. Virtual Reality has already been used in a CSCW system for graphical design [24]. But cooperative work often involves collaborative writing or the sharing and modifying of textual information. Several computer systems offer the collaborators a common workspace in which to sketch, write out, and structure ideas [11]. Our interface provides such a space in three dimensions. The whole field of collaborative writing for technical reports and documentation is a potential area for three-dimensional structure mapping. The advantages of such mapping are that all the collaborators can visualize and agree upon the evolving structure of the document and that each collaborator can see the relationship of his or her contribution to the whole. Three dimensional concept maps should provide a clearer view of this evolving structure. Also, this interface could be used in the teaching of technical and scientific writing. Hypertext systems are now being applied to the teaching of writing [12, 16]. The advantage is that the fluid associative structures of hypertext are good for the teaching students to brainstorm and organize ideas. Again concept maps help here: students can better understand that a document needs to have a clear structure, and adding the third dimensions should help to clarify structure.

INTERFACE ISSUES

A three-dimensional textual interface raises all the issues associated with constructive hypertext, as well as some issues peculiar to Virtual Reality. A key question is how to represent and interact with both hierarchical and associative
(linked) structures. Even in two-dimensional hypertext systems, there is a potential conflict between the representation of hierarchy (if this is permitted) and the representation of association. How can these two kinds of relationships coexist in three dimensions? Furthermore, in a constructive hypertext the user is allowed to edit and add verbal text as well as structure. How will the text be represented within a three-dimensional space, so that the user can both read and change it with relative ease? The user must also be able to create, modify, and delete links. Yet another question is whether an immersive or Virtual Reality interface will improve spatial orientation and navigation through the structure. Finally there are problems associated with the current state of Virtual Reality hardware. Can text be legible and changeable with the resolution of current VR displays? Will a helmet-mounted display or goggles add an unacceptable burden for the user? Let us consider some of these issues in greater detail.

Representation of hypertextual structure
In converting concept maps from two dimensions to three, we can employ any one of a number of mappings. We can use the third dimension to represent subordination, as is done with the cone trees of the PARC system. We can devise more complex spatial structures to represent hierarchy, such as the metaphor of the building, with floors and rooms. Here hierarchy is indicated by incorporation: a building consists of floors, which in turn consist of rooms. As in two-dimensional maps, special typed links can be used to indicate hierarchy (parent-child relationship). Ideally, the mapping should be adaptive, so that each user can make his or her own representations of hierarchy and association.

In systems with two-dimensional concept maps (e.g. Storyspace), the user can manipulate the structure by selecting and dragging nodes with the mouse. The analogous interaction is possible in three dimensions. However, grasping structures in space proves to be difficult with the current generation of equipment, mostly due to the small range and the noisiness of the trackers. Smart selection mechanisms (e.g. gravity or prediction algorithms) may help; multimodal input can also aid in selection [20]. Furthermore, new tools will have to be developed for multiple selection. The interaction and selection mechanisms of course depend on the mapping of the hypertext as well. For example, the building metaphor makes it hard to select a group of nodes that are scattered around in the building.

Presentation and interaction with verbal text is more obvious in two dimensions than in three. The reason is that alphabetic writing has always been conceived as two-dimensional and is normally arranged in linear rows that form paragraphs. In a 3D interface, the text itself is an object and must be located in the space along with other objects. One possibility would be to incorporate the traditional page into the 3D space; pages of text would float near the annotation objects to which they are associated. Another strategy would be to wrap texts around objects. Still another would be to exploit the architecture of the virtual space itself by displaying text on walls or other surfaces. Finally, text could be animated in 3D space. On the other hand, the problems of displaying text in 3D space could be avoided rather than solved. That is, we could use conventional 2D windows and mice to display and interact with the text, while using 3D views (and special devices) to interact with the hypertextual structure.

Linking in Three Dimensions
The essence of hypertext is linking, and the representation of links, both in the text itself and in concept maps, is a fundamental feature of any system. Standard methods in two-dimensions are to employ some link marker (asterisk, footnote marker, special symbol) in the text and to draw an arrow in the concept map. Three-dimensional hypertext offers special opportunities and special problems in link representation. Again a line or arrow can be used, but other solutions are possible, including iconic representations of doors, windows, and tunnels. In principle, any icon that indicates motion or transition from one place to another can be used to represent a link, and the range of possible icons is greater in three dimensions than in two. Some three-dimensional representations can convey more information than their two-dimensional counterparts. For example, a link that is represented as a door or window offers the user a preview of the content of the target node, because the user can look through the window or doorway. Types of links can also be represented by animation as well as color and shape: three dimensions permit an additional degree of freedom in animation.

One motivation for representing hypertexts in three dimensions is that two-dimensional concept maps can become illegible because of many intersecting links. When the third dimension is used, the link lines are no longer all coplanar. It may be possible to improve legibility further by exploiting a non-Euclidean coordinate system; that is, by creating fish-eye views and other "distortions" of the space.

Following links is an essential part of browsing hypertextual documents. In a three-dimensional interface, activating links can be done by selection as in a two-dimensional interface. But the interface might also permit movement through the link (door or window), "shooting" at the link, or other techniques appropriate to the particular link representation. Many conventional hypertext applications move in a discrete jump from source to destination node. The heightened spatial perception in three dimensions may make such jumps disorienting; continuous movement, as in driving or walking, from source to destination may be more effective. Exploiting the driving metaphor, the interface could represent a path as a vehicle that moves slowly between the links. The user could decide to enter or exit this "tour" at any point.

Navigation
In word processing, navigation is well defined: the arrow keys and scrollbars allow linear movement through the information. Searches constitute jumps from one point in the
linear text to another. Hypertext expands navigation by giving the user the ability to follow links. Three-dimensional representations can both restrict and broaden methods for moving through the information space. The technique of flying is a common navigational tool in 3D systems. Looking and flying with the helmet-mounted display facilitate browsing as well as goal-directed navigation. Global and detailed views can be obtained just as in the real world by drawing near to or moving away from structures. On the other hand, an important constraint in the use of flying and other techniques is the need to maintain user's spatial orientation and sense of place. Flying at high speeds, the user can get lost or pass out of the virtual world entirely. This problem can be minimized by limiting the navigable space, for example by constructing a boundary cube. An overview map that indicates the user's position can also help to maintain orientation [4].

The 3D hypertextual view can be browsed by following links. Other techniques include text searches and filters. Instead of moving the user from one place to another, such tools could rearrange and change the structure of the document by highlighting occurrences, hiding non-relevant clusters, and so on. The PARC system includes "gardening" tools to "prune" or "grow" its cone trees [21, p. 5]. When multiple users have access to the same structure, however, such rearranging can cause confusion, unless some sort of protocol is established for distinguishing between private views and the public space.

Hardware
Exploring and writing information spaces with current helmet-mounted displays can only be experimental: it is unrealistic to expect that people will be willing to wear a helmet for this application. Furthermore, the current hardware makes it difficult to read and interact with large amounts of text. Ordinary CRT displays have much better resolution, and resolution is clearly an important factor for applications that process text. (With the visual acuity of the current helmet-mounted display, the user is legally blind. Of course, the hardware continues to improve and diversify. For example, a different technology, using stacked monochrome LEDs and a pivoting mirror, can already provide a resolution of 720 by 280 pixels, which should make text more legible [19]. We can also expect improvements in helmet-mounted displays with color LEDs.) Furthermore, a 3D cursor is not a very accurate pointing device for selecting text, and the use of a physical keyboard is almost impossible for a helmeted user. Constructing a virtual keyboard in the user's visual space is not a realistic solution either, because the user would have to type while wearing one or two bulky gloves. A better current solution, we believe, is to use conventional 2D windows and a mouse to interact with the words. Research continues on the use of two-dimensional input devices such as the mouse to manipulate objects in 3D environments [10].

Another important question is whether a tactile interface is appropriate for abstract information. The history of writing in recent centuries has been one of increased abstraction and distancing of the author from the text: through the printing press, the typewriter, and now the word processor and hypertext [2]. Does the user want to touch and hold blocks of text as he or she does other objects in virtual reality? Pen-based computing would suggest a return to the tactile quality of handwriting, but we have yet to see how and where pen-based computing will be used. In the long run voice recognition may supplement or supplant keyboard and pen-based input -- for word processing and hypertext as well as for menu-driven interfaces. So voice recognition might also be a long-term solution for generating text in 3D spaces. The user would dictate the text, which would appear floating in space and could then be move and edited as needed.

THE WORLD PROCESSOR PROTOTYPE
Our prototype, called the World Processor, manipulates both structure and text in three dimensions. The World Processor constitutes a first attempt at solving some of the interface issues outlined above. In general, our design philosophy has been to adopt the "obvious" extrapolations from the two-dimensional hypertextual concept map in order to test the effectiveness of these extrapolations in three dimensions. To facilitate testing, we use a standard hypertext application to enter text and structure offline. The World Processor then imports two-dimensional hypertext and locates it in a 3D space. The system allows the user to browse through and to modify this spatialized hypertext. The system also includes a very simple text editor and several widgets for creating new elements and moving existing ones, so that the user can build hypertexts from scratch.

Our Virtual Reality platform consists of a Silicon Graphics Indigo Elan workstation, a Virtual Research helmet-mounted display, a dual-receiver Ascension Bird 3D tracker system, and some video-equipment to convert the graphics to NTSC format. One of the Bird receivers, which measure both position and orientation, is mounted on the helmet-mounted display for updating the viewpoint within the virtual scene. Another receiver is mounted on a conventional mouse with a piece of velcro. The receiver transforms the mouse into a three-dimensional pointing device, and its position and orientation are coupled to a small virtual object ("cursor").

For quick rendering of the three-dimensional scenes, we use the gl library for the Silicon graphics machine. A C-library has been written to permit the construction of and interaction with virtual environments. This library, the Simple Virtual Environments (SVE) library, supports loading and saving of objects (hierarchical groups), initializing special devices, and so on. The library also contains an event callback mechanism with default event handlers to make it easier to develop and modify the user-interaction. Each object in a scene consists of multicolored polygons, polylines, and texts; each object is stored in a separate ASCII-based object file. A scene file lists all the instantiations of objects and their properties.
Importing, Navigating, and Modifying Structure in the World Processor

Storyspace, a hypertext authoring system written for the Apple Macintosh, is used to create hypertexts for conversion. Storyspace supports concept maps in two dimensions, and it also allows hierarchical grouping of the nodes. The conversion routine parses the Storyspace document, maps the information to a virtual environment, and saves the environment as a SVE scene. A built-in previewer shows the converted Storyspace document in a small 3D window; the viewpoint can be changed with the conventional mouse and keyboard. In the current system the mapping method is determined at compile time by a small source file that links the storyspace elements and their properties to a visual representation.

The current mapping locates hierarchy beyond each cell (Figure 2); however, other mappings are possible (e.g., rooms or cone trees). In the future, the user interface will include support a "mapping-editor" that enables the user to alter the mapping of the hypertext document interactively, similar to the glyph binder that manages the mapping of datasets to their visual representation in scientific visualization.

In the World Processor, each information node is represented by an object located in space, called an annotation. Objects may take any shape that can be constructed using the system's graphic primitives: typical representations are cubes, spheres, or text symbols used as icons (e.g., a quotation mark or question mark). Textual labels or titles can also be attached to objects. Links currently are rendered as white lines connecting the source and destination annotations. The entire structure of links and nodes can float in a void, or there can be an arbitrary background to enhance visibility. The background can even be occupied by a VR model. For example the structure may annotate the model of a room, building, or a city, and the relationship between the annotation structure and the background model may be significant.

As in many other VR applications, the World Processor allows the user to "fly" around in space. During the flight (as long as the right mousebutton is pressed) the flight direction is determined by the orientation of the cursor. Thus, by pointing down and pressing the right mousebutton, the user will descend in the virtual environment. If the button is pressed for a relatively long time, an acceleration mechanism will automatically increase the flying speed. The acceleration variable is reset when the user stops flying. The prototype also offers a way to visualize a path that connects several annotations by means of flying vehicles.
Creating and Modifying Text in the World Processor

To display text, the user first selects an annotation object; to select the object, the user points and shoots (as in a video game). The left mousebutton initiates this action, and the direction of the bullet is determined by the cursor. The point-and-shoot technique means that the user does not have to move close to an object in order to read its contents and thus can activate multiple annotations without navigating. Furthermore, shooting engages the user/reader more than the commonly used point-and-grab metaphor. Two modes of displaying text have been implemented. The first is the handheld page. After pointing and shooting, the user finds the text displayed on a block attached to the cursor. The user can then move the text block by moving the cursor (just as we can move a piece of paper back and forth in front of our eyes). The user does not have to approach the annotation object itself in order to read the text. The second method involves simple animation of the text. Characters are extruded from the annotation block to create a marquee effect. One of the function keys is used as a toggle switch between the two modes.

In order to edit the text of a selected annotation, the user presses the 'home' key on the keyboard. A page (colored rectangle) appears floating in space near the node, and text is displayed on the page (Figure 5).

Various auditory and visual cues aid the user in exploring and manipulating structure. The auditory cues reenforce visual cues to confirm user-initiated events, such as touching an object or starting to fly. Visual feedback alone is used to indicate a continuing status: for example, whether the cursor is in continuing contact with an object (the object will have a different color) or whether the user is flying (a yellow arrow will extend from the cursor, Figure 4).

CONCLUSION

As we indicated above, the design philosophy for the World Processor is to extrapolate the hypertext interface from two dimensions to three. Preliminary usability tests, involving creative writers from a local arts group in Atlanta, seem to
show ease of use, consistency of the interface, and sufficient functionality to build complex information spaces. In the first exploratory sessions, the subjects found the interface compelling and continued to create annotation objects and texts for more than twelve hours. However, many refinements are needed to make the interface convenient for less dedicated users. Wholly new interface concepts may evolve in the coming years. We have already mentioned speech recognition as a future solution; we have also been experimenting with voice annotation as a means of putting verbal information (though not ASCII text) into 3D models [25].

We believe that both interface designers and users will come to recognize the importance of representing text and other semantic objects in three-dimensional spaces. Increased computational power is making high quality graphics available on workstations and personal computers, and even Virtual Reality need not be prohibitive expensive [19]. We believe that users will begin to feel an increasing need to locate text in the graphic space. The desktop and windows metaphors of current desktop computing will likely be supplemented or replaced with various three-dimensional metaphors. But these new metaphors will still require text for labels, names of objects, and other kinds of communication. Such interfaces themselves constitute concept maps of their contents. Hypertext and hypertext-like applications are also becoming more common. Collaborative writing is becoming an important application for hypertext, and a World Processor should be a good place for writers to fashion and edit shared documents. Three-dimensional hypertext may also be a part of the next generation of authoring systems for computer-based training and education [14].

REFERENCES


