Designing with a 2 $^{1}/_{2}$D Attitude

Colin Ware
Data Visualization Research Lab
Center for Coastal and Ocean Mapping
University of New Hampshire
Chase Ocean Engineering Lab
24 Colovos Road
Durham, NH 03824
Phone (603) 862-1138
Fax (603) 862-0839

Computer graphics gives us freedom to design highly interactive 3D information spaces, but when will these be better? An analysis of human space perception reveals that the dimension of egocentric space towards and away from an observer is perceived very differently from the dimensions orthogonal to the line of sight. Because of the this property of perceptual space a strategy of designing with a 2 1/2D attitude is advocated and elaborated in a set of design principles.

Introduction

Driven by the computer game industry, most personal computers are now capable of displaying complex 3D virtual worlds, complete with realistic textures on all surfaces. This raises the possibility that we should be designing 3D interactive information spaces. But will a computer interface looking like a 3D video game actually be more useful?

One way of addressing the 3D versus 2D is to apply more than a century of research into the psychology of space perception. This tells us that "3D" is not a simple unified concept in perceptual terms. We can have degrees of three dimensionality, using different spatial "depth cues" and we can choose to increase or decrease the amount and type of 3D information. This article applies findings from depth cue research and object perception theory to address the issues of when we should use 3D and how much 3D we should use. This analysis leads to the conclusion that the 3D information provided should depend on the application and the task. Even when a rich set of 3D cues are justified the three spatial dimensions should not be treated at though they were homogenous – different design rules should be applied in the z direction (in and out of the screen). This is called designing with a 21/2D attitude.

There are physical reasons why going from 2D to 3D in a display does not add nearly as much information as, say, going from 1D to 2D. A useful way to think about human vision is in terms of Gibson's concept of the ambient optic array. This describes the information available to an observer as a bundle of rays that converges on a particular viewpoint in space as illustrated in Figure 1. The ambient light array defines a fundamental physical constraint on vision: we can only perceive one package of color information for each ray in the bundle. What can be presented on a monitor, or any conceivable computer display is, at best, a perfect simulation of the ambient optical array

and with the exception of depth of focus information, this can be done quite effectively with a flat screen.
To increase the information available in the optic array the brain can integrate information from the two eyes, giving stereoscopic depth, and integrate information from a moving scene, or rotating objects to give kinetic depth. However, even when this information is added there is much less information available in the z direction than in the plane orthogonal to the line of sight.

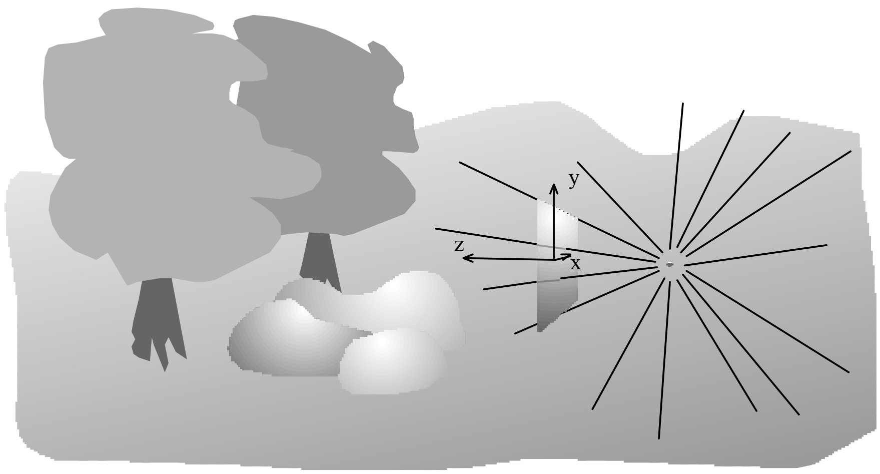


Figure 1. An ambient optical array defines the information at a point of observation. A window defines the information that can be represented on a flat screen.

By looking at how the visual system processes spatial information we can discover which kinds of 3D information can help us perform different computer-based tasks. Much of the research into space perception is couched in a framework of "depth cues". Depth cues are different kinds of optical information that we can use to understand spatial layout. We can apply depth cue theory directly to the problems of computer-based work. Which depth cues are important depends on the task being performed. Depth cues are often divided into two classes. The pictorial depth cues that are preserved in a static image such as a photograph , and the other non-pictorial cues. We begin with the pictorial depth cues.

Occlusion

Overlap or *occlusion* is one of the most important depth cues. The term occlusion refers to the fact that objects near to us, block or *occlude* objects further away. An object that occludes another almost always appears closer. Occlusion over-rides all other depth cues except under very special exceptional circumstances. Figure 2a illustrates how occlusion leads to the impression of layering in depth. Occlusion is an ordinal depth cue, it only tells us that one thing is closer than another, not by how much. This can be useful; for example, the tabbed cards illustrated in Figure 2b use occlusion to provides rank order information in addition to rapid access to individual cards. Although modern graphical user interfaces (GUIs) are usually described as being 2D, they are actually 3D in a non-trivial way. Overlapping windows rely on our understanding of occlusion to be effective.

a b

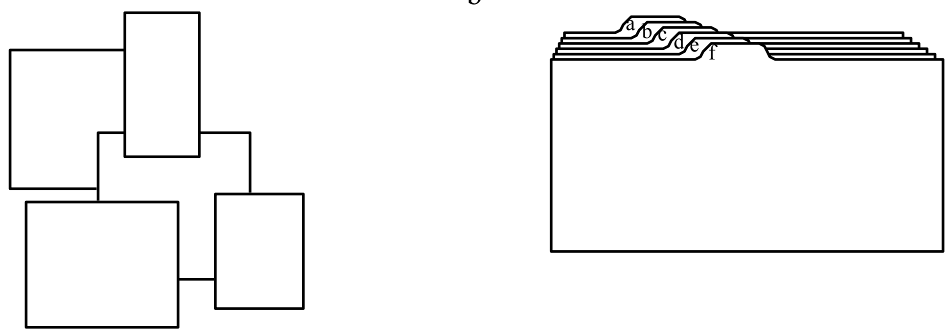


Figure 2. (a) When one object blocks, or “occludes” another it appears closer. Under most circumstances occlusion over-rides the other depth cues. This basic depth cue is essential to modern window-based interfaces. (b) Careful use of occusion enables a small tabs provide acess to a much larger objects.

Cast shadows and shape-from-shading

Cast shadows can be used to enhance the apparent distance between overlapping objects as shown in Figure 3. Relatively subtle shading effects can give a 3D appearance to buttons and widgets and this is widely used to enhance the perception that certain objects can be grasped and manipulated. This kind of shading of widgets is another instance of a 3D depth cue being used in current GUIs, and simple cast shadow effects are becoming widely used in design graphics.

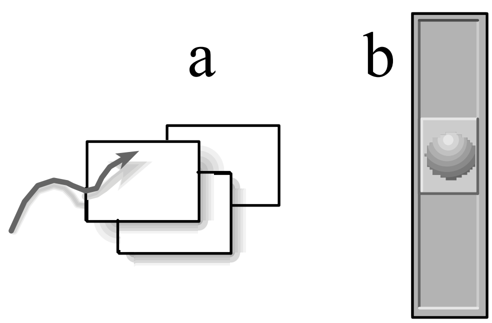


Figure 3. (a) cast shadows provide information about the relative space between objects. (b) shading information provides information about the 3D shape of surfaces. Here it helps make the slider seem more tangible.

Perspective

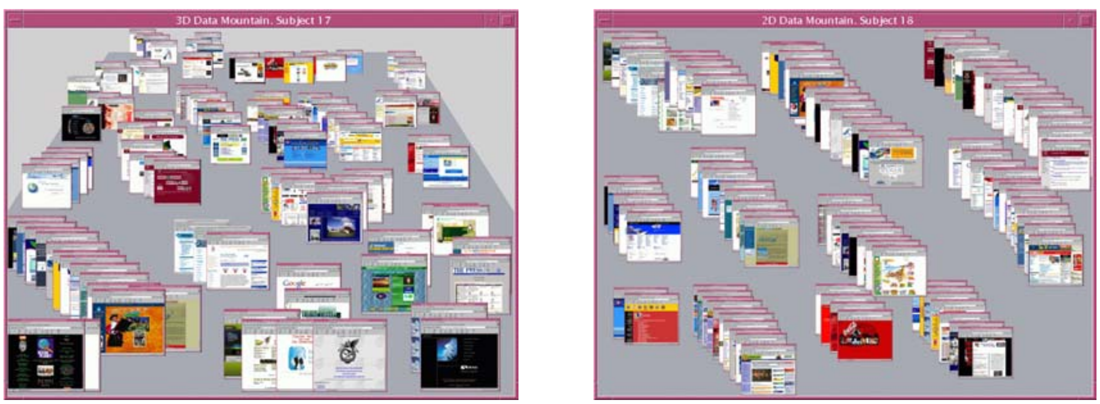
The term perspective covers a set of depth cues that all derive from the basic geometry of mapping a 3D scene onto a flat screen with a particular viewpoint. This geometry gives rise to three related depth cues.

Linear perspective refers to the fact that projections of parallel lines converge on the picture plane.

Texture gradients, refers to the reduction in size and increase in density of texture elements with distance,

Size gradients refers to the way more distant objects are represented as smaller on the picture plane.

The fact that the brain corrects for perspective size changes is called *size constancy*. Size constancy means that we can perceive object sizes more-or-less accurately despite large variations in retinal image size. Thus, to some extent perspective can be viewed as something that must be compensated for, not as anything useful in information display. There is little evidence that a perspective picture lets us see more than a non-perspective image. A study by Cockburn and McKenzie showed that perspective cues added no advantage to a version of Robertson et al's (1918) Data Mountain display. The version shown in Figure 4b was just as effective as Figure 4a. However, both of these version make extensive used of other depth cued (occlusion and height on the picture plane).



a b

Figure 4. Variations on data mountain display Cockburn and McKenzie.

Structure from Motion

It is much easier to see the 3D organization of a rotating cloud of data points than a static cloud of data points and for this reason rotation is a standard feature of molecular modeling packages. The impression of 3D structure that results from a 2D projected image of a 3D arrangement of lines rotating in space is called the kinetic depth effect. The kinetic depth effect comes under the more general category of structure-from-motion cues and these include the moving images that result as we move through the world.

Structure-from-motion cues are generally more important than stereoscopic depth in providing information about complex 3D information structures. Glenn Franck and I evaluated the importance of structure-from-motion and stereoscopic depth in making it possible to see large interconnected nets of information in 3D like that shown in Figure 5. This very large nested graph, shows the structure of digital switch software. It contains more than 30,000 nodes and a similar number of edges, although it is deeply nested so that most of it is not visible at any one time. Our results obtained using simplified patterns showed that a network about 60% larger could be interpreted with a stereoscopic view, but a network 220% larger could be viewed with a structure from motion information (Ware and Franck, 1996).

Structure-from-motion cues are also important in providing the phenomenological impression of spaciousness and this is likely to be important in 3D design applications (Ware et al, 1993). For example, if the goal is to help someone appreciate the spatial layout of a new design for a kitchen or an automobile interior a moving animation will be much more effective than a series of stills.

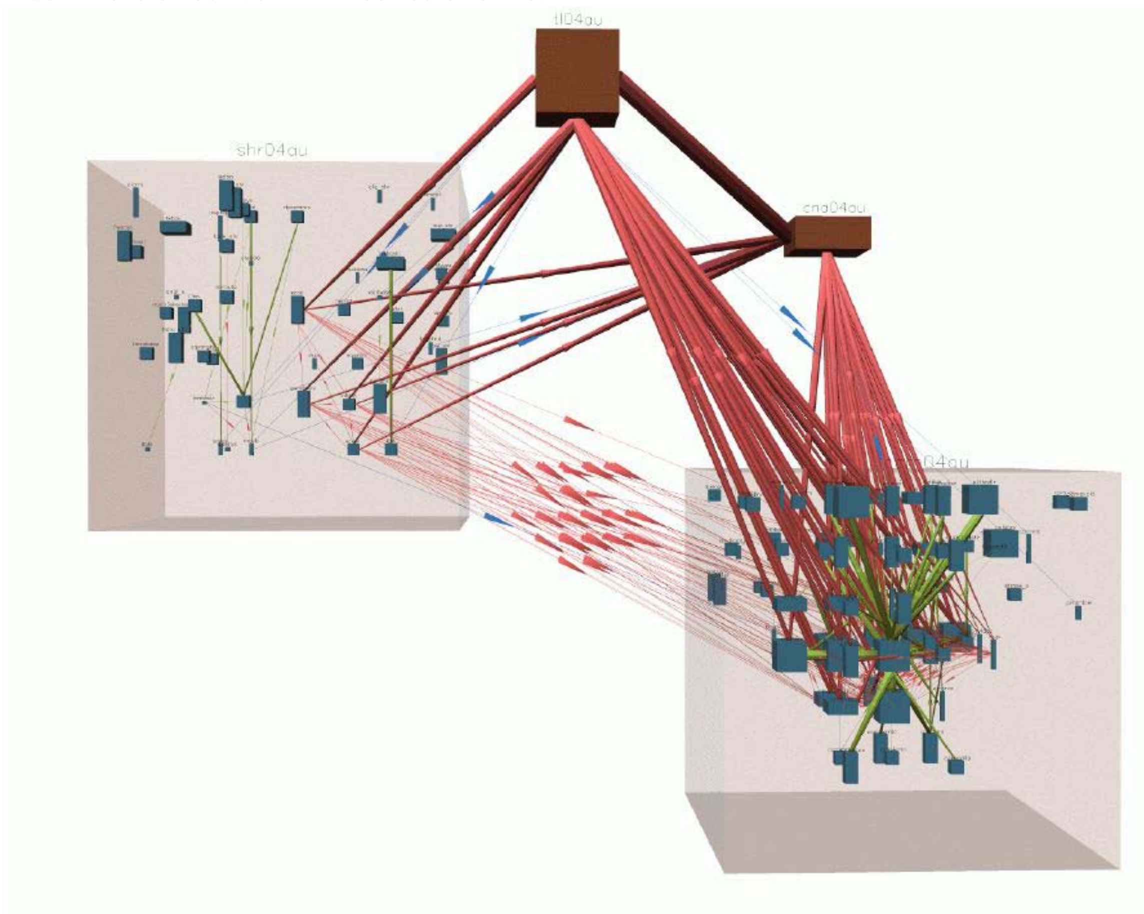


Figure 5 . The architecture of a large digital switch is revealed in a 3D visualization. If the visualization is rotated more structure is revealed. But occlusion is a problem.

Stereo

Stereoscopic depth refers to the fact that the human brain can use differences between the two images in the two eyes to extract information about distances to objects in the environment. To create stereoscopic depth, devices such as shutter glasses provide images (computed with different viewpoints) separately to the two eyes. Such devices are often said to provide “true” 3D. But this is misleading advertising. In fact stereoscopic depth, is generally one of the less important depth cues for most tasks. As already mentioned, structure-from-motion is stronger than stereo and so is occlusion. The fact that occlusion overrides stereoscopic depth can be a problem when trying to create a virtual object lying in front of a monitor using a stereo viewing system; the apparent occlusion of the object by the frame of the screen causes the stereoscopic depth effect to disappear.

In real-world interaction stereoscopic depth is mostly important in carrying out fine manipulation tasks, on objects held close to the body. Thus as we develop computer based environments for fine design work stereopsis will become a more desirable depth cue to provide.

Other cues
There are a number of depth cues that are normally thought to be less important than the ones listed above. These include, depth of focus information and the convergence of the two eyes. Depth of focus cannot properly be simulated with a flat screen display and this is a significant problem as we push the limits of virtual reality displays.

3D perceptual space is not uniform

The preceeding analysis shows that 3D perceptual space has a very non-uniform structure. However, it is not only the near-far dimension that is distinct. The other two dimensions also have special properties that are important and relevant to interaction design. If we define a coordinate system centered on the observer's viewpoint as illustrated in Figure 1. We can make the following general observations about the different dimensions.

Y-axis. (Up-down) The Y vertical dimension is defined by gravity. It has an "up" polarity. We generally move around the world on a roughly horizontal ground surface and are much less free to navigate in the up-down direction. To make a virtual data space familiar is often desirable to maintain an up direction in a data space so that the up vector is aligned with the up on the screen display.
X-axis. (Left-right) Turning to the left or right, along with progressing forward are the main characteristics of terrestrial locomotion. Children and some adults are often confused by left right symmetries, Presumably because of the left-right symmetries of our bodies.
Z-axis (Near –far). The z direction is also the direction of normal locomotion. We mostly navigate the world by moving forward in the direction we are facing. A far lower information density is available in the Z direction in comparison with the other two.

As we navigate through unfamiliar environments we gradually build up a mental map. However, the process is slow and the map is often highly inaccurate. Not surprisingly, people can build a mental map much more quickly if they can see the entire set of objects they are trying to map objects from a single vantage point. Information spaces that consist of a set of rooms with corridors between will be hard to mentally map. A more open 3D environment that with an interface allowing the navigator to zoom out and get an overview will be easier.

3D object perception

Thus far we have mostly been considering how people perceive the relative positions of objects in space, but the theory of how we perceive individual 3D objects can also contribute to the 3D versus 2D debate. Object perception happens in a series of stages. First simple features are extracted in a massively parallel processing system, next the visual scene is segmented into regions, by contour, color and texture, and finally 3D object structures are identified. This process is illustrated in a very simplified form in Figure 6. Early stage feature processing and segmentation is all 2D and the transition to 3D happens at an intermediate level famously characterized as a 21/2D sketch by David Marr (Marr and Nishihara, 1978). The 21/2D sketch is a kind of enhanced feature map containing information about the contours of objects, rough distances to objects and the orientation of object surfaces. From the 21/2 D sketch the visual system extracts 3D structures. This process involves the selection of inflection points in the objects to find where structural components meet. The tangent lines of the object silhouettes are assumed to be surface tangents and the brain infers rounded surfaces. In an elaboration of Marr's theory, Biederman (1987) proposed that 3D primitives are identified from the low level information. He called these geometric primitives "geons" and suggested that they consist of a set of generalized cones and ellipsoids. A structural skeleton linking the geons makes up the object topology. Surface color and texture are secondary characteristics of geon objects.

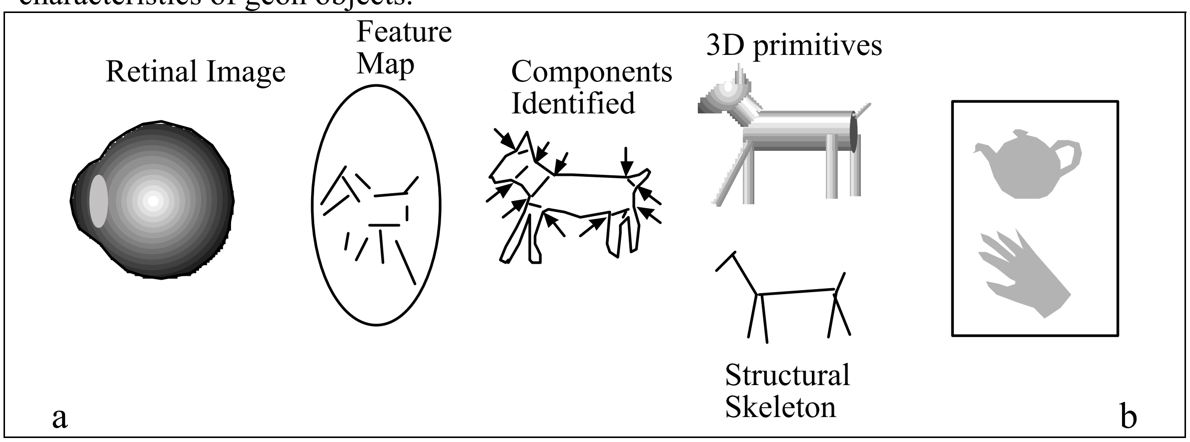


Figure 6. (a) Stages in the perception of structured objects. (b) canonical views of objects make them easier to identify.

The theory of object perception has a number of design implications for 3D diagrams. Components will be easier to identify if they are laid out in the x,y plane making it easier to perceive the silhouette and hence find the structural skeleton. This also means that certain "canonical" views of objects will be easier to identify as Figure 6b illustrates. The idea that there are 3D primitives for perception of objects suggests that we should use 3D elements for diagrams in preference to 2D shapes. A study that Pourang Irani and I recently carried out showed that diagrams like the one shown in Figure 7 are better than conventional UML diagrams in two respects. They make it easier to identify sub-components and they are easier to remember.

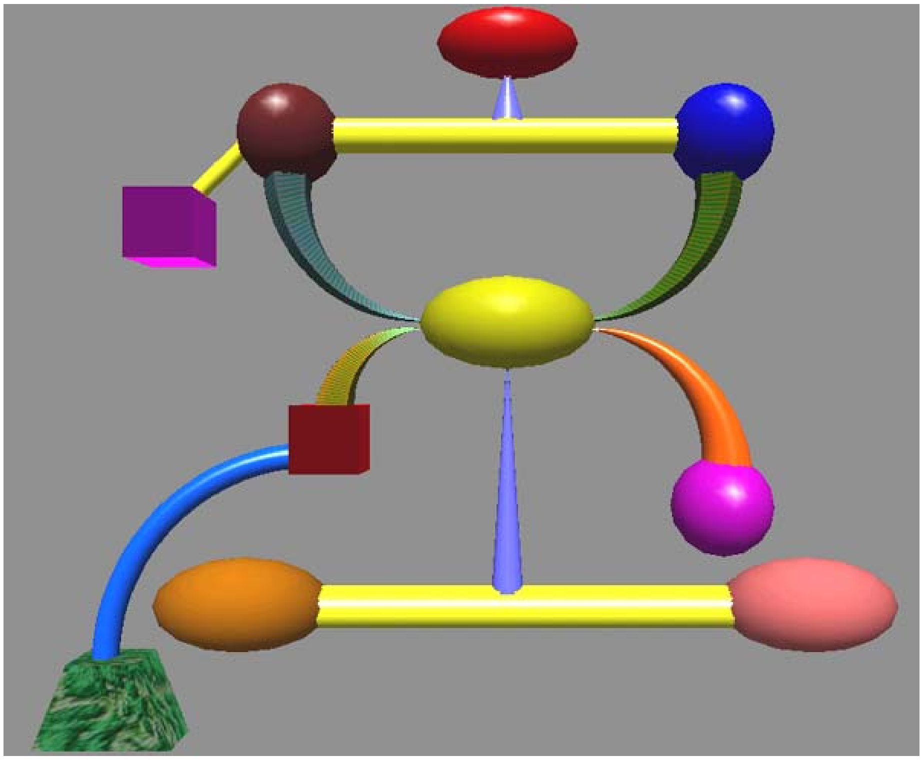


Figure 7. It is easier to remember and analyze diagrams with 3D components.

The 21/2 D Design Attitude

A way of summarizing the implications of human spatial perception is that we should design with a 21/2D attitude. We should not treat display space as a homogenous volume since we cannot see beyond the first object that blocks our line of sight in any direction, but nevertheless, the judicial use of depth cues and make information displays more effective. The 21/2D design attitude we propose here is inspired to some degree by David Marr's 21/2D sketch but there is no one-to-one correspondence with his theory. What follows is a set of design principles that make up the 21/2D design attitude.

1) Use 3D objects to represent data entities. Objects should be constructed in 3D but, especially if they are used as symbols they should be arranged to have the most canonical 2D view. The 3D object perception theories of Marr and Biederman suggest that if we make use of 3D objects in user interfaces, these objects will be easier to identify and remember than simple 2D shapes. Also, data structures will be understood better if they are mapped to object structure because the human visual system will automatically extract the object structure (and hence the data structure) as part of normal perception.

2) Emphasize 2D layout. Where object structures are important the design should be such that this structure is principally laid out in the 2D plane, especially if a static view is required. It is especially important to design with careful used of occlusion to make all objects visually accessible. Text should normally be displayed in the image plane for legibility.

3) Use depth cues selectively as needed. The fact that the depth cues can be characterized separately to one another means that we can use them separately. 3D is not an all or nothing choice. When it is important to understand complex structures in 3D use structure-from-motion depth cues Shape-from shading is important in helping us identify 3D objects. Motion parallax and stereoscopic cues can be used to make larger 3D structures perceivable, if there is sufficient empty space between the elements. Linear perspective, texture and size gradients are only to be used if a 3D workspace is defined as in Figure 4.

4) Use 2D layout support navigation in 3D spaces. 3D information mazes are generally to be avoided since it takes people a long time to generate a mental map of such a space. If a 3D virtual world is used it should be constructed so that navigation cues are always visible on the screen. In practice this means that the user should not have to turn around to see some particular exit. Doors that are out of sight behind the user are less useful than more artificial navigation devices that are always visible. Where a large data space is to be viewed an overview map should be provided or, it should be easily possible to easily zoom out to get overview. Overview maps will be easier to use if they have the same general perspective as the magnified image (another 2D constraint), since people find it difficult to cognitively integrate data seen from radically different viewpoints.

Where does the 2 1/2D design attitude apply? Certainly not to all information design problems. Since text is a fundamentally 2D medium and because fonts can be designed more legibly with a flat screen in mind, there would seem to be no good reason for displaying text in 3D. Even when text labels are required for 3D objects they should normally be made so that they are parallel to the plane of the screen. Thus for the set of tasks associated with text: reading it, entering it and editing 2D will be more appropriate..

At the other end of the spectrum are tasks where the information to be viewed is fundamentally 3D. This covers applications such as interior design, animated 3D character design, viewing molecular models or the topography of another planet, and the aesthetic design of a 3D artifacts. For these cases adding as many depth cues as is practical would seem to be the right approach. Although there are tradeoffs, for example using structure from motion cues incurs a high computation cost and stereo viewing involves specialized equipment. Eventually it will be possible to dynamically model depth of focus effect as someone looks around the scene, and although this is beyond what is feasible now, it will add to our appreciation of virtual 3D spaces.

It is in the large middle ground of information display that a 21/2D design attitude is likely to be most useful. This includes the design of the interface that makes files, and application programs easily accessible as well as new applications for managing and interpreting data networks, data bases and complex software systems. Increasingly visualization is becoming important in analyzing and managing these large complex information systems. A 21/2D design attitude that uses 3D depth selectively and pays special attention to 2D layout would seem to provide the best match with the limited 3D capabilities of the human visual system.

References

Biederman, I. (1987) Recognition-by-components: A theory of human image understanding, Psychological Review, 94(2) 115-147.

Cockburn, A., and McKenzie, B. 3D or not 3D? Evaluating the effect of the third dimension in a document management system. ACM CHI2001 proceedings 434-441.

Gibson, J.J. (1986) The ecological approach to visual perception. Lawrence Erlbaum Associates. Hillsday, New Jersey.

Irani, P and Ware, C. (2000) Diagrams based on structural object perception. Conference on Advanced Visual Languages. Procedings, Palermo, Italy, May.

Marr, D., and Nishihara, H.K. (1978) Representation and recognition of the spatial organization of three-dimensional shapes. Proceedings of the Royal Society of London, B. 200, 269-294.

Robertson, G., Czerwinski, M., Larson, K., Robbins, C., Thiel, D., and van Dantzich, M. Data Mountain: Using spatial memory for document management. In Proceedings of ACM UIST'98 153-162.

Ware, C. and Franck, G. (1996) Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions. ACM Transactions on Graphics.15(2) 121-139.

Ware, C, Franck, G., Parkhi, M. and Dudley, T. (1997) Layout for Visualizing Large Software Structures in 3D. In Proceedings of Visual 97, conference on visual information systems. San Diego, December, 215-223.