1997

# A Hybrid Immersive / Non-Immersive Virtual Environment Workstation

N96-057
Department of the Navy

Report Number 97253



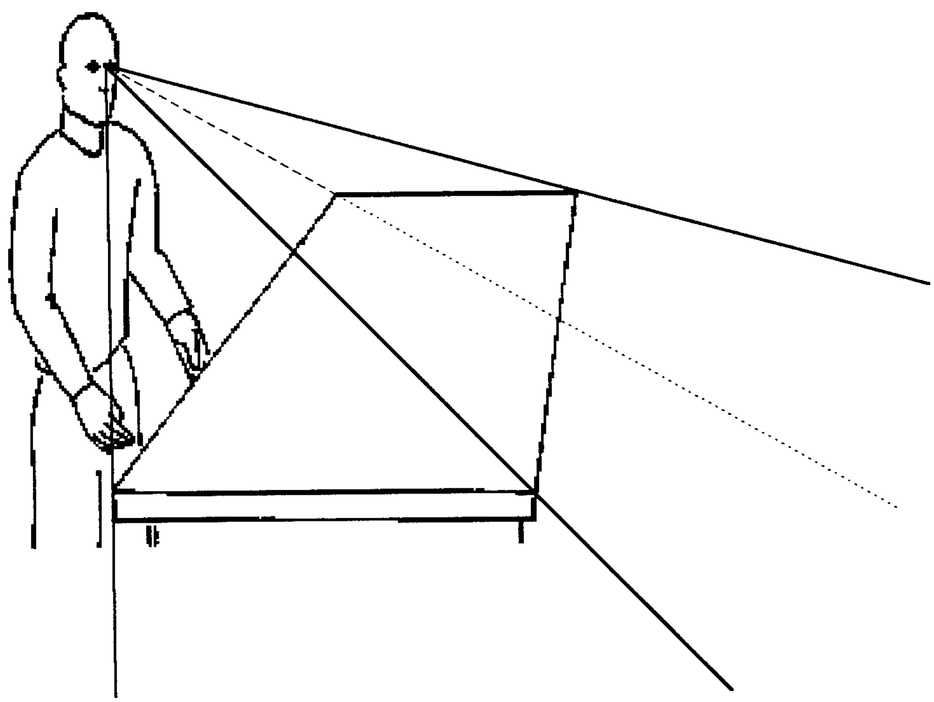
**Submitted by:**
Fakespace, Inc.
241 Polaris Ave.
Mountain View, CA 94043
Phone (650) 688-1940
Fax (650) 688-1949

## Introduction

The accurate representation of maps and other symbology by the virtual environment is required so that decisions based on insights gained from using the system can be made with confidence. The presentation of virtual environments with immersive display technologies such as head mounted displays have been studied in some detail as evidenced by the research done by NASA, UNC, MIT and other institutions. The evaluation of Non-Immersive stereoscopic interfaces is less well developed as the technology is relatively new.

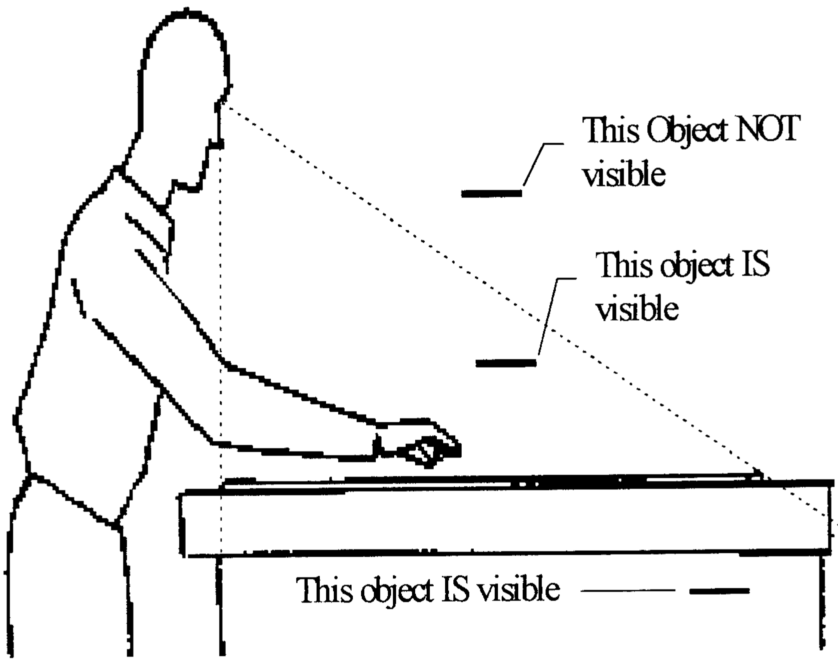
## The Non-Immersive Display Environment

The Non-Immersive interface to the system we have been working on projects a stereoscopic image onto a table surface from where it can be seen in stereo by wearing special glasses. The stereo images are formed in a specific region as illustrated below. As you can see, the stereo images are created in a cone from the user's eye to the image surface.



Stereo imagery may be placed in the pyramid which extends from the user's eyes through the corners of the image to infinity.

The virtual images appear both above and below the plane of the image at the table surface. As you can imagine, once an object gets high enough off the table surface, the stereo illusion breaks down. This appears to be based on two factors. Firstly, the visual accommodation and convergence cues significantly conflict leading to visual discomfort. Secondly, the working volume gets smaller as objects are positioned higher off the table surface, therefore when you move, the object may fall outside the working pyramid. The use of icons around the perimeter of the visual space to indicate nearby but not visible objects would be useful so that the area being examined in detail does not lead to the user missing wider range activity.




Visibility is determined by the objects position relative to the viewing cone

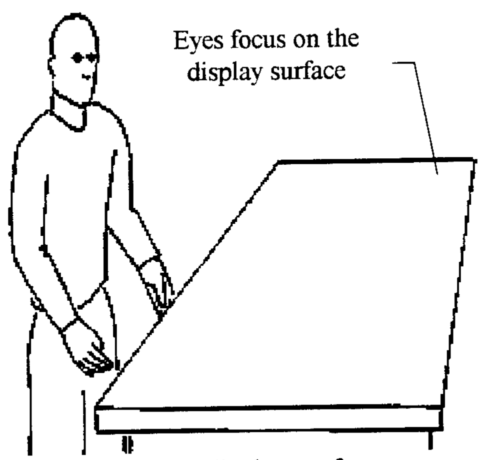
## Non-Immersive Display Visual Characteristics and Artifacts

There are several cues which your eyes use to determine the three dimensional spatial relationships of the objects around you. The relevant cues considered here include: perspective, stereopsis, occlusion, accommodation, and convergence. Other distance cues, for example, haze and level of fine detail are generally more applicable to very distant objects or serve an ancillary role in the determination of the visual scene.

The perspective nature of the display and software mean that objects become visually smaller as they move away from the viewer into the distance. This distance cue is based on a prior knowledge of the relative size of objects. For instance, if two similar sized ships are seen and one appears to be smaller than the other, then it is presumed that the smaller one is further away. This is a stereo cue which works equally well in both a monoscopic and stereoscopic environment. Because perspective is assumed to be available in all systems such as this, it is difficult to gauge it's relative importance. This is primarily because the system makes an implicit assumption that the scene is being drawn with perspective.

Stereopsis is used by your visual system to determine the relative depth of objects on the basis of retinal disparity. Stereopsis is the depth information one derives from the fact that the two eyes see slightly different views of a scene. The different images are presented to your eyes on the Non-Immersive display by using one of several approaches. The most well known approach is to place electronically controlled liquid crystal shutters in front of each eye. These are switched on and off synchronously with a projector that alternately displays left and right images as calculated by the image generator. The accuracy with which these two images are computed depends on several factors, the most critical of which is overall calibration. Getting the stereo pair correct is of great importance as this is perhaps the strongest depth cue for systems such as this when your head is stationary.

By drawing the scene on the Non-Immersive display from each of the user's two eye locations, the correct occlusion cues are also created. Thus, as one moves one's head, closer objects obscure others which are further away. The occlusion cues are particularly useful for discerning objects which do not have a lot of surface detail. These occlusion cues are also related to the small motions which one's head and eyes perform on a subconscious level. You may have noticed people moving their heads slightly back and forth just before they reach out to precisely touch or grab something. The precision with which the occlusion cues are faithfully recreated on the Non-Immersive display is obviously heavily dependent on the system's temporal performance. The lag between a move of the tracker on the user's head and the display of the corresponding image means that the occlusion cues will not be perfect during significant head motion. Occlusion cues are probably as strong as the stereoscopic cues and are particularly strong when there are a number of objects on the table and your head is in motion.

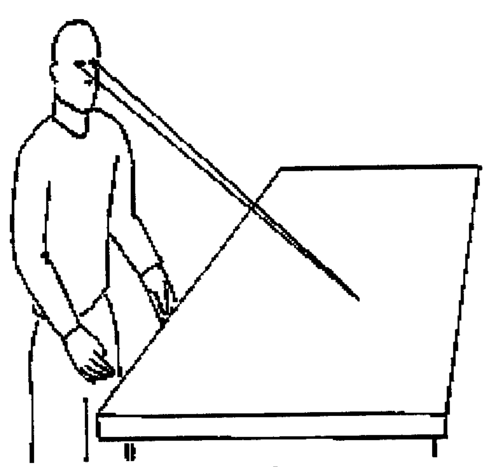



Accommodation on the display surface

The accommodation of your eyes on nearby objects also provides depth cues. Accommodation is the deformation of the eye's lens by the ciliary muscle and associated ligaments. The deformation of the muscle controls the convexity of the lens and thus the focus of the eye. The physiological process by which a focused image is maintained on the retina provides the brain with a measure of the distance to the object being looked at. This is done by reporting the amount of deformation of the lens required to maintain focus. The accommodation stereo cue works in a range of about 7 yards and is a relatively weak cue.

If one focuses an eye on a finger and then move the finger closer to the eye then one will feel the exertion as the eye tries to maintain focus with the ciliary muscle. The stiffness of the crystalline lens increases as one ages so the ability to focus the eye on close objects is gradually diminished as we grow older. In terms of the Non-Immersive display system, the accommodation cues are well matched for objects that are close to the surface of the table. As objects are positioned above the table surface, so the virtual position and the actual position from an accommodation standpoint diverge. The eye will always accommodate on the table surface as that is where the actual image of the objects is drawn. This is true no matter what the actual position in space of the virtual object. This disparity between the accommodated position and the perceived position from stereopsis is not generally a problem as accommodation is not the primary depth cue.

The last depth cueing mechanism we will discuss here is convergence. Convergence is the orienting of the eyes to make the line of sight of each eye meet at a point on an object. As the two eyes work together to asses the position of objects in the real world, they rove over the scene maintaining their relative orientation so that the line of sight of each eye intersects with the object being inspected. The lateral orientation of the eye is controlled by the four recti muscles which rotate the eye around the horizontal and vertical axes. The control mechanism of the brain uses the differential exertion and extension of the horizontal angular muscles (internal and external recti) to gauge the degree to which the two eyes have to angle in towards the nose to look at an object. For example, when looking at a very distant object, the muscles will be relaxed and the line of sight of each eye will be parallel. When looking at a close object (the tip of one's nose is the most obvious example) the eyes are significantly angled inwards. The degree of this tilt is used by the brain as an additional cue to assess the relative distances between objects in the scene. The convergence cue is used by the visual system as a relative measure rather than as an absolute range finder. This cue is most effective for objects closer than 12 feet or so. In the Non-Immersive system these cues are fairly good when objects are close to the table top. This is one key reason for using a surface near horizontal rather than vertical for the Non-Immersive user. The convergence cues are completely incorrect for stereo displays on a flat image plane oriented substantially vertically (such as a normal desktop monitor).



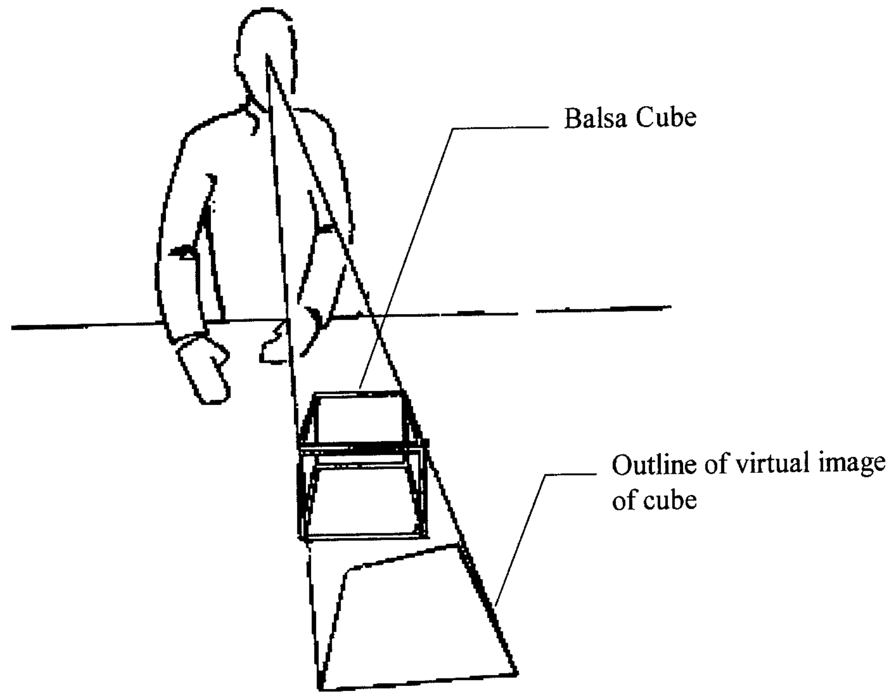
Convergence on the table surface

There are also some psychological assumptions the user makes when using the Non-Immersive display. These assumptions have to do with expectations as to what is going to be seen. Some of these contributory factors are that the user commonly uses their hands to point into the image and also rests their hands on the border around the screen. These physical cues as to where the images are located in space seem to reinforce the notion of a flat image plane with objects in the vicinity of that plane. It may be possible to slightly alter the physical and visual environment of the Non-Immersive display to reinforce alternative visual interpretations and thus achieve a more compelling experience. For example, currently, the edges of the displayed image end abruptly and this creates a very sharp divide between visual objects that are on the table and those that are not. Altering the periphery of the display may alter the perception of the images and the interpretation of a very abrupt border. Making this transition smoother or using the perimeter of the display to show other related information might make the border seem less abrupt.

## Calibration Issues for the Non-Immersive system

As previously described, the overall system calibration issues are important for achieving a high degree of visual accuracy. It is most important that what appears to be true on the Non-Immersive display is, in fact, an accurate representation of the actual situation. In looking at the various physiological aspects of the display, it is evident that the various visual cues need to be produced with a high degree of fidelity. In our work on the calibration issues we have experimented with a couple of different scenarios. For instance, we considered how to extract the calibration data through the use of virtual points that float above the table and having the user move to line them up. This approach and many of the others we considered seemed mathematically intractable and complicated. The most effective method we have found for checking the overall system calibration is to use a real object which lines up with a virtual one.

Wire frame balsa wood shapes which are of a known size and shape were constructed. Corresponding virtual models were built of the same simple size and shapes. The calibration can then be easily verified by starting up the software with the virtual model sitting on the table surface. Then the real physical model can be placed resting on the table. From the user's perspective, the two models should exactly line up. All the combined static errors of the system can then be seen. In our initial tests, the results were pretty good and the misalignment of the real and virtual objects was on the order of 0.25 inches on a 90 inch display. The factors which contribute to this error include: tracking accuracy, accuracy of the offset from the tracker to the eyes of the user, errors in projector alignment, screen flatness, and optical distortions. The temporal characteristics prove harder to quantify. As you move your head around, the computer-generated model lags behind the real one. When you stop moving, the real and virtual models come into alignment. This incongruence is much less noticeable when there is the virtual model only. The mis-registration is far more apparent when there is a real model to continuously compare with while the head is moving.




Calibration with real objects

In addition to these errors in the visual display, the accommodation and convergence cues for the real and virtual models are different. The accommodation for the real model is, of course, correct but the virtual one is not. The accommodation difference is not distracting for objects that are less than about 6" off the table surface. Objects which appear out of this band start to create difficulties.

## *Other visual Effects*

### Flicker

Currently, the stereo images on the Non-Immersive display are achieved by showing alternating left / right images. These images are permitted to arrive at the appropriate eye through the use of shutter glasses. These shutter glasses contain electro-optic filters which are remotely controlled to be semi-transparent or opaque. In the open state the polarizing filters permit about 32% of the light to reach the eye. In the dark state the transmission is about 1/1000th of that. The images are presented from a single CRT based projector at between 96 and 120 images per second. Thus, each eye sees a new image at between 48 and 60 times per second. This is generally sufficient for normal viewing needs. It does have two disadvantages however. Firstly, the glasses are active and thus require batteries. Secondly, since the glasses change state in response to an infrared beacon, the infrared sensor in the glasses requires a clear line of sight to the beacon.

This is not the ideal situation for a deployable solution. Ideally, the glasses would be passive so that they are simple and have no batteries to fail at inopportune moments. This may be achieved by one of several methods including passive stereo and perhaps other novel approaches.

## Passive Stereo Methods

A number of groups have developed various products known as Z-Screens or passive stereo shutters (Tektronics, Stereo Graphics, NuVision, and others). These are shutters that may be placed in front of a randomly polarized source and alternately pass one polarization state or another. These systems can then be viewed with passive glasses which contain the appropriate polarizers for each eye. By using circularly polarized light, it is possible for the system to work even as you tilt your head. However, these systems do introduce a chromatic distortion as you tilt you head. This manifests itself as a color shift to blue in one direction and to brown in the other. This is not particularly desirable but does permit relatively inexpensive and passive glasses. There is an additional good feature of passive systems. One may choose to use two projectors and polarize each projector's output to a different state. The images are simultaneously presented on the screen and are not time multiplexed. This is an advantage since then the imagery is on the screen all the time and has the potential to be twice as bright as a system in which the image time is multiplexed between the two eyes. Another ancillary advantage of this approach is that it creates redundancy in that it uses two independent projectors. If one of these projectors were to fail for some reason, the system would continue to operate in monoscopic mode by taking the glasses off.

## Depth of field and accommodation

Rear projection display systems are typically not very bright at the moment, they must be used in subdued lighting. This means that the eyes' pupils will dilate and the depth of field of the eye will be less than it is during daylight. This, in turn, may exacerbate some of the visual discontinuity between the accommodation and convergence cues. The eye will still accommodate and converge on the table surface as before, but any real object in the scene will require the eye to shift it's focus more than in a brightly lit environment. The magnitude of this effect is hard to quantify at this point.

## Occlusion By Real Objects

It is interesting to see what happens when one starts to use the virtual workspace created by the Non-Immersive display as you might use a real space. If you reach out to point at something, your real hand does appear to point to the virtual object from your point of view. If the system is well calibrated, then pointing at the virtual object will position your hand just as if there was a real model there on the Non-Immersive display and you were pointing at it. Of course, your hand is going to occlude parts of the image. Unfortunately, the occlusion of your hand and the virtual objects will not be strictly correct because the images on the display surface always fall behind your hand.