SPATIAL INPUT/DISPLAY CORRESPONDENCE
IN A STEREOSCOPIC COMPUTER GRAPHIC WORK STATION

Christopher Schmandt

Architecture Machine Group
Massachusetts Institute of Technology

ABSTRACT

An interactive stereoscopic computer graphic workspace is described. A conventional frame store is used for three-dimensional display, with left/right eye views interlaced in video and viewed through PLZT shutter glasses. The video monitor is seen reflected from a half silvered mirror which projects the graphics into a workspace, into which one can reach and manipulate the image directly with a "magic wand". The wand uses a magnetic six degree-of-freedom digitizer. In an alternative configuration, a graphics tablet was placed within the workspace for input intensive tasks.

CR catagories: I.3.2 [Computer Graphics]: Graphics Systems; I.3.6 [Computer Graphics]: Methodology and Techniques; I.4.8 [Image Processing]: Scene Analysis-Stereo.

Key Words: Stereoscopic display, half silvered mirror, three-dimensional digitization.

INTRODUCTION

The last decade has seen a resurgence of interest in three-dimensional imaging, such as holography, commercial 3-D movies, experimental anaglyphic television broadcasts, and lenticular 35mm cameras. Some of this enthusiasm has overflowed into computer graphic applications, most noteably computer aided design, medical imaging, and seismic exploration, but widespread appreciation of the utility of stereoscopic displays has yet to happen. This may be in part due to the lack of convincing interactive capabilities of such displays.

The fields with the most interest in three-dimensional display are those with a need to analyze complex spatially structured real world data, gathered by or part of a process requiring expensive data acquisition hardware (hence the willingness to experiment with relatively expensive displays). Although a variety of techniques exist for stereoscopic display, there has been little work done in interacting with 3-D data by 3-D modes: using motion parallax to look around objects [Fisher, Littlefield], peeling away interfering data layers, or constructing additional data types in image space itself.

Figure 1. A cutaway view of the work station, showing the view through the half silvered mirror of the video monitor and user's hand.

In short, the blossoming of conventional two-dimensional interactive computer graphics has yet to penetrate the three-dimensional world. Such graphic inter-activity requires three components: display technology (e.g., digital frame store), input technology (e.g., tablet, mouse) and algorithms or techniques link-ing the two.

This paper will briefly survey existing three-dimensional display and input technologies. It will then focus on a style of interaction characterized by spatial correspondence between the display and input devices; we desire to build a stereoscopic workspace which will allow a user to reach into the midst of a three-dimensional image and interact with it at exactly that point in space which he is touching.

The workstation display is a conventional color frame buffer using television inter-lace and special opto-electric shutter glasses for stereoscopic viewing. The monitor is viewed through a half-silvered mirror, which projects its image into an illuminated workspace (figure 1). The user views his hand and the graphics, mixed by the mirror, overlapping in three-dimensional space. Input technology has included a six degree-of-freedom magnetic digitizer and a conventional magneto-strictive tablet.

A series of experiments will be discussed. Early experiments exercised the work station, using the magnetic digitizer, in various applications; despite early optimism, numerous limitations on inter-action were noticed. After discussing these, we will explore alternative appli-cations, with more clearly defined inter-action goals, as well as a second work-station configuration which uses two-dimensional tablet input to build 2½-D databases.

DISPLAY TECHNOLOGY

Stereopsis is the perception of depth from the binocular fusion of the slightly off-set views of our left and right eyes. Stereo display systems must somehow cor-rectly generate left and right eye views, and present them separately, avoiding crosstalk, to the respective eyes. As this paper focuses on real-time inter-active computer generated display of information, we will limit this section to various forms of computer controllable video display [Okoshi].

A number of techniques for display of three-dimensional video have been devel-oped. Points may be displayed on a CRT in synchronization with a vibrating mir-ror, which generates a true three-dimensional image under ordinary viewing

Figure 3. The PLZT glasses and electronic controller box.

conditions [Stover, Fuchs]. Two CRTs may be used in a head mounted display, using optics to deliver each image to the respective eye [Sutherland, Callahan]. Two CRTs may be viewed through orthogonal pairs of polarizers, as is commonly done in movie theater projection, although crosstalk may be a problem and the viewer must keep his head fairly erect. Two images may be mixed on a single CRT as a color anaglyph and viewed through red and green/blue filter glasses. Finally, two images may be time-multiplexed on the CRT and viewed through electrically controlled glasses [Roese].

The latter technique was used in our work. Conventional interlaced video actually displays two fields sequentially; one field consists of the even scan lines, the other of the odd lines. Using an ordinary frame buffer, the right eye view is drawn on even lines, with the left eye view on odd lines (figure 2). This mixed image is viewed through wafers of a lead lanthanum zirconate titanium (PLZT) ceramic, a very fast electrically triggered shutter (figure 3). Electronics detect the video vertical interval, alternately opening and closing each eye's shutter while the corresponding field is being displayed. Our controller and PLZT glasses were obtained from Honeywell, with additional glasses from Matsushita.

In terms of the computer graphics algo-rithms, the depth of an object is a func-tion of the distance between the even and odd line portions of the image (figure 4). Zero disparity makes the graphic appear on the surface of the screen, as we are accustomed to viewing. When the horizontal position of left and right eye portions of the object are different, eye conver-gence will make the object appear behind or in front of the screen. This relation-ship can be expressed as:

$$dx = IO \left(\frac{d}{Ds+d} \right)$$

where dx is image disparity, d the apparent distance of the object from the screen, Ds the viewer's distance from the screen, and IO the interocular distance (about 65mm).

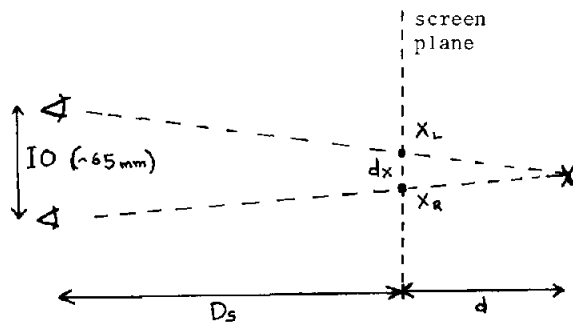Figure 4. Ocular convergence and depth perception. The left eye sees an object at Xl on the screen, the right eye sees it at Xr, with disparity dx. The object will be perceived to lie at distance d behind the screen. Note that if dx were negative the object would appear in front of the screen.

INPUT TECHNOLOGY

Digitizing three dimensional data points for system input has been a more difficult problem than stereo display. One approach uses various forms of mechanical linkages (such as fishing line) connected to measuring devices [Roberts, Clark]. Although high accuracy may be obtained, these linkages limit freedom of movement and are somewhat awkward.

Other techniques use infrared light sources, worn on the body, as transmitters and light sensing hardware as recievers to track multiple points in three space [Burton, Woltring]. Although the commercial product (Selspot) is quite expensive, recent work indicates lower cost tracking may be possible in simplified configurations [Ginsberg]. The main problem with light transmission is obscuration by the user's body.

We chose a magnetic six degree-of-freedom digitizer developed by Polhemus Navigational Sciences [Polhemus, Rabb]; while still fairly expensive, it avoids most problems of previous technologies and is very accurate under ordinary circumstances. This digitizer returns the position (x,y, and z) and attitude (azimuth, elevation, and roll) of a small sensor relative to a larger radiator 40 times per second. Radiator and sensor electronics are connected to a Nova 1200 minicompuer which is

interfaced to our laboratory's machines (Perkin-Elmer 3230's).

Three mutually orthogonal dipole coils in the radiator generate a rotating magnetic field, which induces currents in the sensor, a plastic cube about 1.5 cm on edge which also contains a similar array of receiver coils (figure 5). The sensor is small and light, connected to nearby preamplifiers by a thin, flexible cable. As position is sensed magnetically, there are no awkward linkages, and a user's body does not interfere with the device (although, as will be discussed below, color monitors do).

For the purposes of this work station, a sensor has been mounted at the end of a small wooden "magic wand" which also has a momentary contact bush button switch placed within the handle. The wand, switch, and cable are black, while a white spot on the sensor aids visibility and indicates the "active" portion of the wand. The radiator is mounted out of sight behind the work area. In later versions of the workspace, a conventional magneto-strictive graphics tablet, also painted black with a white spot on the puck, was used for input.
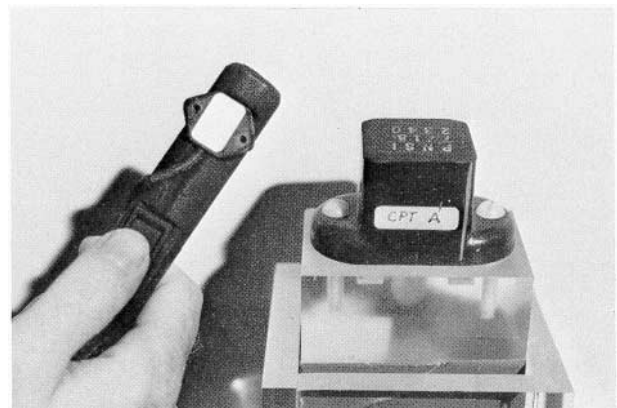

Figure 5. The Polhemus radiator and wand, with the sensor mounted near its end.

THE WORKSPACE

The goal in workspace design was to allow a style of interaction in which spatial correspondence between the particular input and output devices could be maintained. We wished to be able to project the stereoscopic image into an unobscured space which would allow one to reach into the projection with the wand and modify the graphic by touching where it appeared to be. Proper illumination would be essential to maintain the optical mix of the projected image and the user's hand.

Prior work, both at MIT and elsewhere, used half-silvered mirrors to mix computer

graphics with reality. The various head
mounted displays attempted to place
graphics in known positions in the environ-
ment with mirrors mounted immediately in
front of the viewer's eyes [Vickers,
Callahan]. Larger mirrors had been used
to overlay graphics onto keyboards and
tablets [Knowlton, Lippman]. Our task was
to extend this concept into three-space.

A 19 inch color CRT is mounted at a 45
degree angle above and in front of the
operator's head, and viewed through the
half-silvered mirror parallel to the floor
(figure 6). Thus the reflected graphics
appear to lie in the space in front of
the operator and below the mirror. This
space below the mirror is painted black
and illuminated, so he can see his hand in
the midst of the image. In some configur-
ations this lighting was dimmed under
computer control.

The following sections will describe
several early applications for which this
workspace was used, with the wand as input.
Despite the optimism of our early reports
[Schmandt], various limitations interfered
with fully satisfactory interaction.
After discussing these flaws, and our
attempts to remedy them, we will present
a second generation of applications which
matched the capability of this configura-
tion more closely. More recent work has
explored alternative input schemes, using
a graphics tablet.

IMAGE CREATION

3-D Paint

Early work was aimed at developing spa-
tial correspondence algorithms in the
context of a 3-D paint program. Menu
buttons at the front edge of the work
station allowed color selection; when the
user pressed the button in the wand,
colored paint would appear in space, at
the location of the white spot on the end
of the wand (figure 7), dribbling out as
the wand was moved. If the positions of
the graphic and wand were well correlated,
no cursor would be necessary, as the white
spot on the wand is itself a real world
"cursor".

Several calculations determine the posi-
tion of the paint. First, the digitizer's
location, in its own coordinate system,
must be rotated, translated, and scaled
into the screen coordinate system
(actually, that of the projected image of
the screen). This transformation matrix
is a function of the user's point of view,
so it was derived from a calibration step
at the beginning of a session. Next, the
line of sight from the eye to the dig-
itizer is extended to intersect the pro-
jected screen. This gives an (x,y)
coordinate in screen space (that of the

frame buffer, except that x is mirrored).
The z coordinate is used to determine the
left/right disparity with which to draw
the graphic at that point.

In practice, this correspondence works
excellently most of the time and the paint
actually appears to flow from the end of
the wand, provided it is not moved
quickly (see below). One can draw lines,
spirals, knots, etc.; nearer objects
obscure more distant ones, but in turn
draw behind still closer lines.

Depth perception adjuncts

Early in this stage we discovered that
only an experienced viewer could reliably
and comfortably see the proper depth in
the PLZT video image. Purely binocular
convergence is missing many strong depth
cues such as shadows, textures, perspec-
tive, and motion parallax. Of course,
hidden surface obscuration provides an
obvious depth perception aid, but further
cues ease perception greatly.

In all demonstrations, hidden surface
removal is done in real time during image
generation using false color and arithmetic
frame buffer write operations. Associated
with each of the 64 discrete depth levels
are several slots in the color table, one
for each of the eight displayable colors.
Distance is encoded in the high order bits
of color, increasing farther away from the
viewer; while the low order bits determine
which particular color is displayed.
Graphics are added to the screen by writ-
ing the appropriate value (a function of
color and distance) into the frame buffer
under a minimum function; the lesser
(i.e., closer) of the new value and current
value gets written into the display.

When additional depth cues are provided,
depth perception is significantly facili-
tated. The brightness and size of lines
decrease as they are drawn farther away,
as exaggerated perspective and lighting
effects. We noted that although stereo-
scopic depth perception was improved with
these adjuncts, they did not work well
enough in isolation to render the conver-
gence cues redundant; rather convergence,
obscuration, luminance and size all worked
together to give a strong three-dimensional
sense.

VLSI design station

These results were encouraging enough to
move toward a more practical demonstration,
using the wand to input vertices of a 2½
dimensional polygon database. Such data-
bases were later used to generate anima-
tion on a write-once optical videodisc,
for viewpoint dependent image work
[Fisher]. For a practical application, we
chose to work with the graphic component

of a VLSI integrated circuit design station.

Integrated circuits consist of approximately rectilinear components ordered spatially in a laminar manner, with much information contained in how such layers overlap spatially. This content lends itself well to the 2-½ D environment that is suited to real-time frame buffer work. As we were concerned with only the graphic component, VLSI designs of varying complexity were selected from textbooks, and entered by several students using the workspace and wand.

INTERACTION LIMITATIONS

After initial enthusiasm over the ability to accurately track the wand graphically in three space wore off, we were confronted with several aspects of the workspace which detracted from performance in the chosen applications.

Magnetic interference

The most disturbing aspect of this hardware configuration is magnetic interference with the Polhemus digitizer caused by the color CRT. Magnetic fields from the monitor disturb the digitizer in two ways: the normally Cartesian digitizer coordinate space is badly warped, and noise is introduced.

As the magnetic space warping is temporally constant, it was sufficient to derive a single mapping from the warped to a Cartesian three space. This was done by hand mapping magnetic contour lines in the workspace, and approximating them by circular sections. Such an approach works reasonably well except at extremes of the active graphical volume. The unwarped coordinates are then transformed, in software, by a transformation matrix that is generated before each session by having the user touch a succession of displayed dots.

Digitizer noise is induced by strong magnetic field components at the 15.75 kHz horizontal video frequency. Despite the addition of a hardware notch filter to attenuate this frequency in the preamplifier, noise remains excessive with the close proximity of monitor and radiator in the workspace. We placed an ordinary aluminum window screen across the face of the monitor (it can be seen in some of the photographs), which reduces noise to a level such that software filtering could generate useful data.

This additional filtering has the unfortunate effect of introducing a lag, or lack of responsiveness, into the system, with the obvious loss of interactivity. We partially compensated for this with an adaptive filter, which uses a larger

sample window when the sensor is nearer the CRT, and hence more noise prone, but remain disappointed with response time.

Depth judgement

As already mentioned, exaggerated drop off in brightness and premature foreshortening were early additions to aid depth perception, an indication that purely binocular depth cues may be insufficient. Although this enhances perception of the relative distance of pieces of the image, it remains very demanding to properly determine the distance of one's hand relative to the image. Accurate distance judgement is crucial in an input task in which discrete objects may be logically required to lie on coincident planes, or when relative depth between planes conveys needed information.

A partial explanation of this difficulty is that although image components may obscure each other, they are always visable through the user's hand. We attempted, not particularly successfully, to alleviate this problem by allowing computer control of the workspace lighting. When the wand was moved farther away, the lights were dimmed to allow nearer, and hence brighter, components of the image to obscure the hand; unfortunately, the light dimmer response time was poor and the overall effect novel but unconvincing.

Drawing difficulties

Most users of the VLSI design station complained of the difficulty experienced in trying to draw planar objects floating in space because of the lack of arm support and limited hand stability. It was nearly impossible to repeatably draw rectilinear objects with proper right angled corners, and required great care to insure that all vertices of a polygon lay on the same z plane, aggravated by the depth perception difficulty.

Several improvements were tried. An invisible grid of variable resolution is used to force input points to land on lower resolution (x,y) coordinates, making input of orthogonal line segments easier. Subsequent vertices were automatically forced to lie in the plane of the initial vertex, a usefual aid only if the first vertex is indeed at the proper depth. Although these additions are in fact helpful, most users find complex input tasks difficult and tedious.

NEW DIRECTIONS

From the work with VLSI design graphics, several new directions emerged due to recognition of the above constraints. User response had indicated that although the spatial correspondence between graphics and the wand was a simple and intuitive mapping, the arrangment was unsuitable for

tasks highly oriented toward the input of data.

One follow on approach explored the usefulness of the same work station configuration for tasks more oriented toward interaction with or manipulation of previously input or generated data. The other path focused on overcoming input limitations using a graphics tablet instead of the 3-D digitizer. Both approaches retained the stereoscopic video display viewed through the half-silvered mirror.

Data manipulation

Efficient graphical input requires a stable and responsive digitizing technology; these requirements were mutually exclusive in the workspace, with the wand so close to the monitor. Accuracy could be traded off for speed by modifying the software filter, suggesting use of the wand as a target selection, rather than positioning, device.

In the case of complex 3-D data, one may well need to manipulate or highlight the data in order to understand it; interactive graphics should assist this. One of the obvious complexities with a three-dimensional database is the variety of ways objects can obscure each other. Tools are needed to take scenes apart, rearrange them, or accent local volumes for more detailed analysis or separation.

A fairly ambitious implementation of these concepts was attempted in a "Model Kit" project [Nisselson], which allows the rearrangement of 3-D scenes composed of pieces of small, digitized pictures of architectural landmarks, medical images, or microscopy slides as primitives (figure 9). Using the wand in the same work station, the user can move or remove objects, call up new primitives from the photographic database, add text, and save periodic "snapshots" of work in progress.

Many prior limitations of the wand as an input device are avoided by this approach. As it is used much more often to "acquire" one target from many in the scene, highly accurate positioning is less crucial, and response is quickened by tolerating more noise with less filtering.

Tablet input

An alternative input technology was needed for such tasks as building the polygonal components of VLSI design. We mounted a conventional graphics tablet, painted black to avoid image crosstalk, inside the workstation at a 45 degree angle to the floor, such that the image of the CRT screen coincided with the plane of the tablet (figure 10). We desired to retain as much spatial correspondence as possible, with an intuitive 3-D interaction.

The tablet is used to input data only on the plane defined by its surface. Access to the complete visual volume is provided by moving the entire image towards or away from the viewer, under control of a soft "slider" selection graphic, until the appropriate image slice corresonds to the tablet plane. The whole image moves in and out such that one is always working on a level currently displayed with zero disparity (figure 11).

Moving the entire image in the Z direction is done by shifting the horizontal position of the odd scan line graphics with respect to the even lines. This is done by two methods. The odd scan lines can be read out of the frame buffer, then written back with the appropriate new x offset. A coherence table indicating the left and rightmost extent of each line is maintained to minimize the pixel data transfer necessary.

A faster method utilizes two complete frame buffers, one for the left eye view and the other for the right eye view. The two images are mixed into one signal in the video domain by the same electronics that detect interlace to control the glasses; the vertical interval detector selects alternate fields from two video sources. With this configuration, the image can be moved very quickly simply by a horizontal pan by one of the frame buffers.

In addition to greatly increasing the speed with which the image can be shifted in and out, this technique can allow faster image generation and manipulation. Instead of breaking the graphics into an even and odd line portion, preventing use of conventional internal frame buffer primitives (such as "line", "rectangle", etc.), each frame buffer maintains a separate complete image. Left and right eye views can be generated in parallel using the processors in the frame stores.

Tablet input has been used in a variety of styles. The original VLSI design task lends itself quite well to this method, allowing easy input of rectilinear graphic components while retaining needed stereoscopic viewing. Other work will use the tablet to input vertices of 3-D solids, built and edited as wire frame models. Once the final design is satisfactory, slower shaded surface algorithms generate solid model views of these objects for animations. The higher speed, high resolution, and noise immune tablet data is much more suited to these graphically detailed tasks than the wand.

CONCLUSIONS

A stereoscopic video display lends itself well to interactive computer graphics, literally opening up a new dimension of graphical control. Elegant input techniques require careful thought to provide the tools of real use, however.

All work described in this paper exploits the spatial mapping of the graphic display into or over the work area, an arrangement which was found to provide intuitive access to the third dimension. This was without a doubt the most exciting element of the project. Our work certainly suggests that the input configuration and support software will need to be developed around requirements of each individual application, and indeed a task area may require several configurations, some for input and others for image manipulation.

REFERENCES

Burton, R.P. and Sutherland, I.E. Twinkle box: a three-dimensional computer input device. NCC 1974, pp. 513-520.

Callahan, M.A. A 3-D display head set for personalized computer, M.S. Thesis, Dept. of Arch., MIT 1983.

Clark, J.H. Designing surfaces in 3-D. Comm. ACM 19,8 (Aug. 1976) pp. 454-460.

Fisher, S.S. Viewpoint dependent imaging: an interactive stereoscopic display, M.S. Thesis, Dept. of Arch., MIT 1981

Fuchs, H., Pizer, S.M., Tsai, L.C., Bloomberg, S.H., and Heinz, E. R., Adding a true 3-D display to a raster graphics system. IEEE Computer Graphics and Applications, Sept. 1982. pp. 73-78.

Ginsberg, C.M. and Maxwell, D. Graphical marionette. Proceedings, SIGGRAPH/SIGART interdisciplinary workshop, ACM. April, 1983 pp. 172-179.

Knowlton, K.C., Computer displays optically superimposed on input devices. Bell Syst. Tech. J. 56:3 (1977).

Lippman, A. And seeing through your hand. Proceedings of the SID, 22:2 (1981) pp. 103-107.

Littlefield, R.J., Stereo and motion in the display of 3-D scattergrams. Proc., Engineering Society of Detroit Computer Graphics Conference, 1982.

Okoshi, T. Three Dimensional Imaging Techniques. Academic Press, N.Y. 1976.

Nisselson, J. A model kit: a system for constructing three-dimensional interactive graphic models. M.S. thesis, Dept. of Arch., MIT 1983.

Polhemus Navigational Sciences, Inc. P.O. Box A, Essex Junction, VT.

Rabb, F.H., Blood, E.B., Steiner, T.O., and Jones, H.R. Magnetic position and orientation tracking system. IEEE Trans. on Aerospace and Electronic Systems Sept. 1979. pp 709-718.

Roberts, L.G. The Lincoln wand. MIT Lincoln Laboratories Report, Lexington, MA. 1966.

Roese, J.A. and McCleary, L.E., Stereoscopic computer graphics for simulation and modeling, Proc. SIGGRAPH 13:2 (1979). pp. 41-47.

Schmandt, C.M., Interactive three-dimensional computer space, Proc., SPIE conference on processing and display of three-dimensional data, Aug. 1982 (in publication).

Vickers, D.L. Sorcerer's apprentice: headmounted display and wand. Ph.D. thesis University of Utah, Dept. Elec. Eng., 1974

Woltring, H.J. and Marsolais, E.B. Optoelectric (Selspot) gait measurement in two- and three- dimensional space, a preliminary report. Bulletin of Prosthetics Research 17:2 (fall, 1980). pp. 46-52
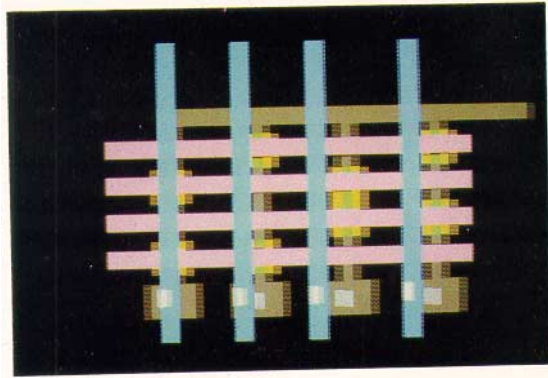
Figure 2. A VLSI design. Note the use of interlace to convey disparity along edges of vertical components.
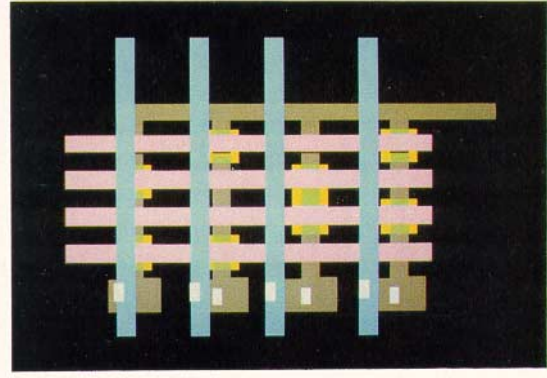

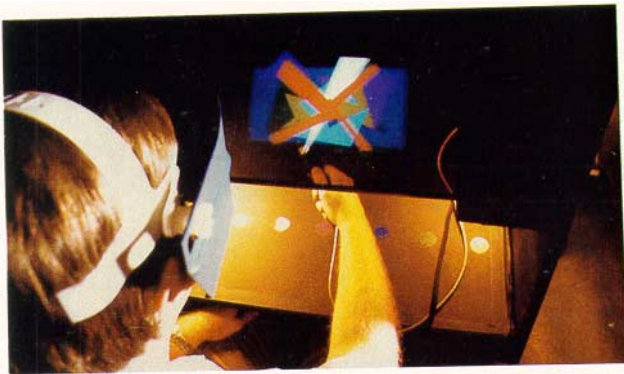Figure 8. One of the single eye views of the adjacent VLSI component.

Figure 6. The work station. The graphics are reflected from a monitor at the same height as the user's head.
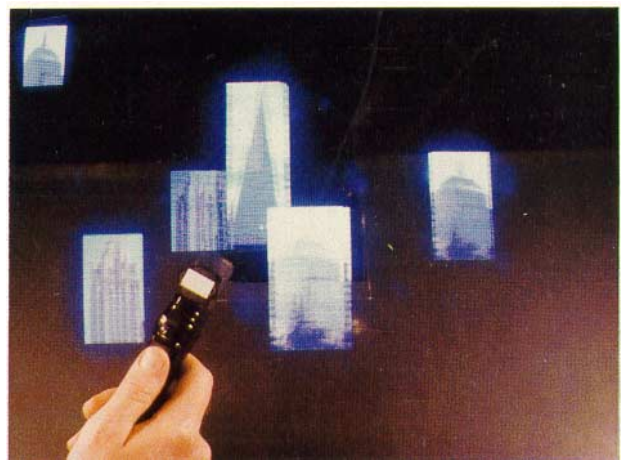

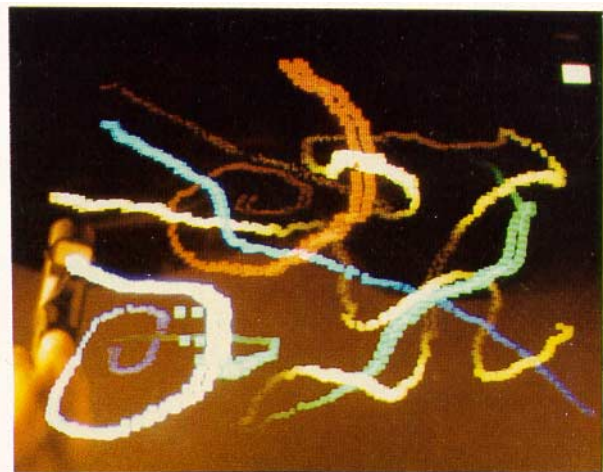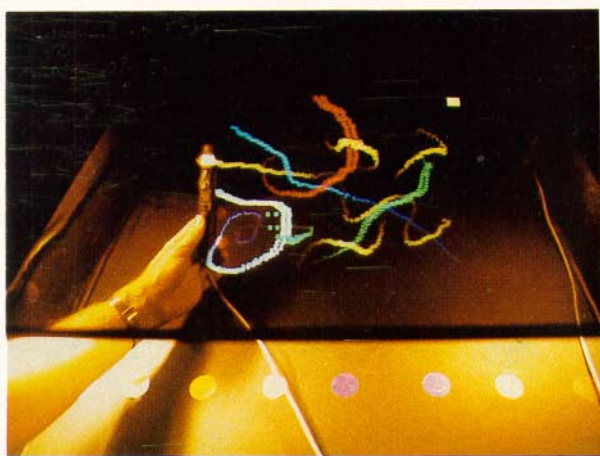Figure 9. Moving architectural icons with the "Model Kit".

Figure 7. Two views through the mirror, showing the overlap of graphics and the hand/wand. Lines that appear double show the disparity between the left/right eye views (presented on alternate video scan lines).
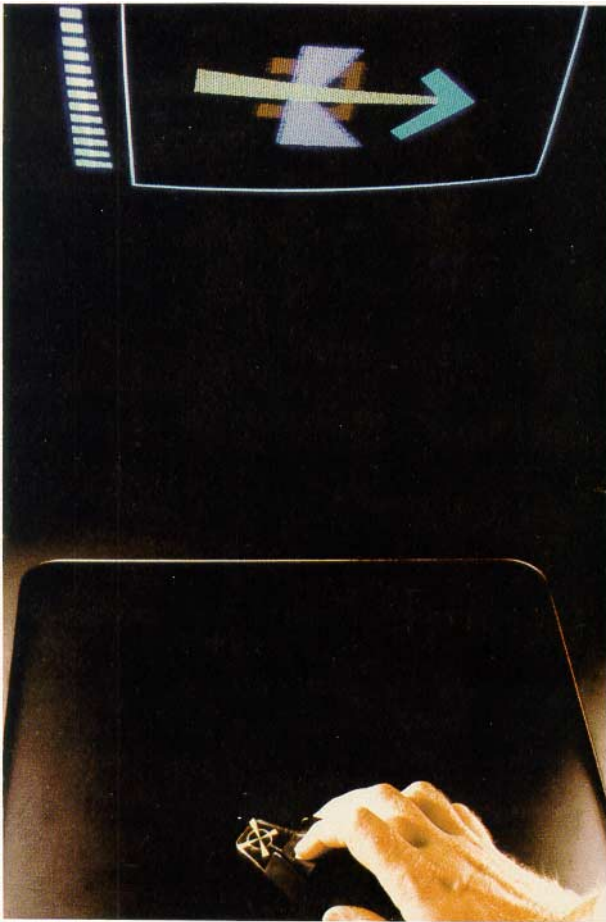
Figure 10.   (Left and above) The tablet workspace configuration.
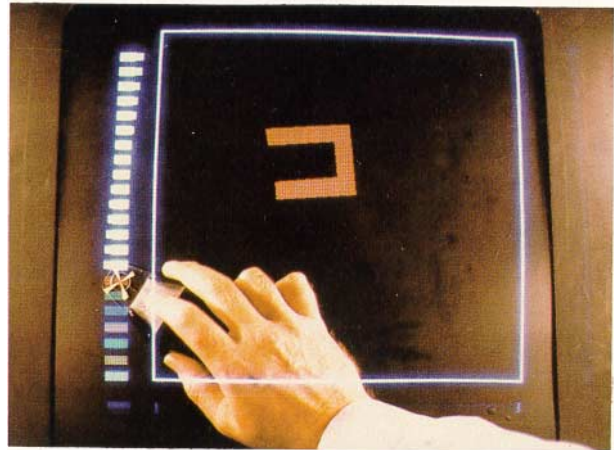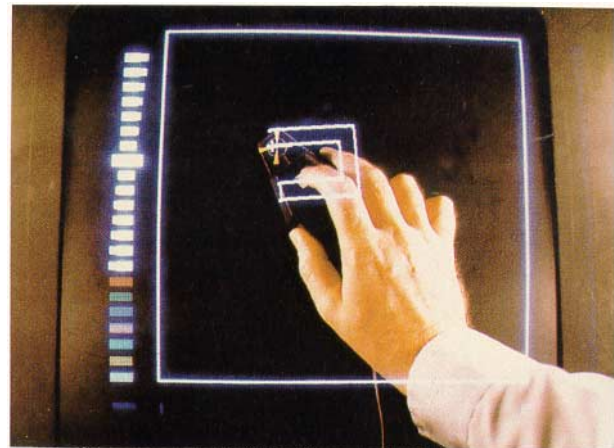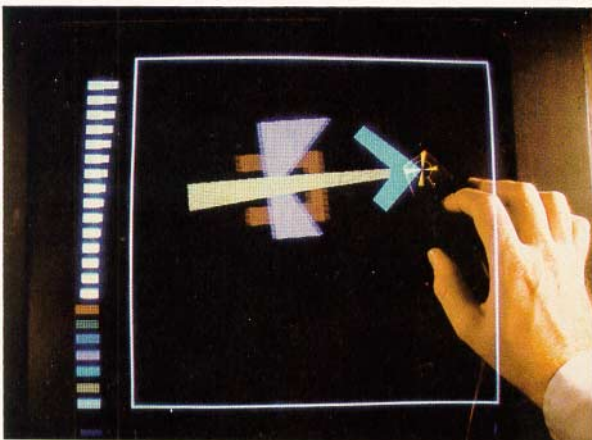

Figure 11.   (Above, and to the left). Use of the tablet to input data.   Rubber band outlines are drawn, and the polygon is filled when a color is selected.   The white rectangles to the left, above the color menu, select the Z plane on which to work.