Proc. of Fall Joint Computer Conference 1968

Historical paper written by Ivan Sutherland, who designed and built the first head-mounted display in the 1960s.

A head-mounted three dimensional display*

by IVAN E. SUTHERLAND**

The University of Utah
Salt Lake City, Utah

INTRODUCTION

The fundamental idea behind the three-dimensional display is to present the user with a perspective image which changes as he moves. The retinal image of the real objects which we see is, after all, only two-dimensional. Thus if we can place suitable two-dimensional images on the observer's retinas, we can create the illusion that he is seeing a three-dimensional object. Although stereo presentation is important to the three-dimensional illusion, it is less important than the change that takes place in the image when the observer moves his head. The image presented by the three-dimensional display must change in exactly the way that the image of a real object would change for similar motions of the user's head. Psychologists have long known that moving perspective images appear strikingly three-dimensional even without stereo presentation; the three-dimensional display described in this paper depends heavily on this "kinetic depth effect."[1]

In this project we are not making any effort to measure rotation of the eyeball. Because it is very difficult to measure eye rotation, we are fortunate that the perspective picture presented need not be changed as the user moves his eyes to concentrate on whatever part of the picture he chooses. The perspective picture presented need only be changed when he moves his head. In fact, we measure only the position and orientation of the optical system fastened to the user's head. Because the optical system determines the virtual screen position and the user's point of view, the position and orientation of the optical system define which perspective view is appropriate.

Our objective in this project has been to surround the user with displayed three-dimensional information. Because we use a homogeneous coordinate representation,[2,3] we can display objects which appear to be close to the user or which appear to be infinitely far away. We can display objects beside the user or behind him which will become visible to him if he turns around. The user is able to move his head three feet off axis in any direction to get a better view of nearby objects. He can turn completely around and can tilt his head up or down thirty or forty degrees. The objects displayed appear to hang in the space all around the user.

The desire to surround a user with information has forced us to solve the "windowing" problem. The "clipping divider" hardware we have built eliminates those portions of lines behind the observer or outside of his field of view. It also performs the division necessary to obtain a true perspective view. The clipping divider can perform the clipping computations for any line in about 10 microseconds, or about as fast as a modern high-performance display can paint lines on a CRT. The clipping divider is described in detail in a separate paper[4] in this issue. Because the clipping divider permits dynamic perspective display of three-dimensional drawings and arbitrary magnification of two-dimensional drawings, we feel that it is the most significant result of this research to date.

In order to make truly realistic pictures of solid three-dimensional objects, it is necessary to solve the "hidden line problem." Although it is easy to compute the perspective positions of all parts of a complex object, it is difficult to compute which portions of one object are hidden by another object. Of the software solutions now available,[2,5–10] only the MAGI[9] and the Warnock[10] approaches seem to have potential as eventual real-time solutions for reasonably com-

*The work reported in this paper was performed at Harvard University, supported in part by the Advanced Research Projects Agency (ARPA) of the Department of Defense under contract SD 265, in part by the Office of Naval Research under contract ONR 1866(16), and in part by a long standing agreement between Bell Telephone Laboratories and the Harvard Computation Laboratory. The early work at the MIT Lincoln Laboratory was also supported by ARPA.

**Formerly of Harvard University

a datum, it clears its input flag. When it has completed its operation, it presents the answer on output lines and sets an output flag to signal that data is ready. In some cases the unit will commence the next task before its output datum has been taken. If so, it will pause in the new computation if it would have to destroy its output datum in order to proceed. Orderly flow of information through the system is ensured because the output flag of each unit serves as the input flag of the next. The average rate of the full system is approximately the average rate of the slowest unit. Which unit is slowest depends on the data being processed. The design average rate is about 10 microseconds per line.

The computer in this system is used only to process the head-position sensor information once per frame, and to contain and manipulate the three-dimensional drawing. No available general-purpose computer would be fast enough to become intimately involved in the perspective computations required for dynamic perspective display. A display channel processor serves to fetch from memory the drawing data required to recompute and refresh the CRT picture. The channel processor can be "configured" in many ways so that it is also possible to use the matrix multiplier and clipping divider independently. For example, the matrix multiplier can be used in a direct memory-to-memory mode which adds appreciably to the arithmetic capability of the computer to which it is attached. For two-dimensional presentations it is also possible to bypass the matrix multiplier and provide direct input to the clipping divider and display. These facilities were essential for debugging the various units independently.

Presenting images to the user

The special headset which the user of the three-dimensional display wears is shown in Figure 2. The optical system in this headset magnifies the pictures on each of two tiny cathode ray tubes to present a virtual image about eighteen inches in front of each of the user's eyes. Each virtual image is roughly the size of a conventional CRT display. The user has a 40 degree field of view of the synthetic information displayed on the miniature cathode ray tubes. Half-silvered mirrors in the prisms through which the user looks allow him to see both the images from the cathode ray tubes and objects in the room simultaneously. Thus displayed material can be made either to hang disembodied in space or to coincide with maps, desk tops, walls, or the keys of a typewriter.

The miniature cathode ray tubes mounted on the optical system form a picture about one half of an inch square. Because they have a nominal six tenths mil spot size, the resolution of the virtual image seen by the user is about equivalent to that available in standard

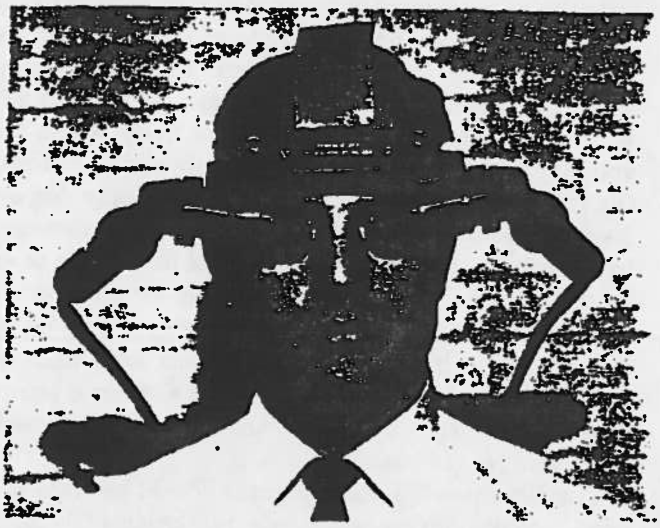

FIGURE 2—The head-mounted display optics with miniature CRT's

large-tube displays. Each cathode ray tube is mounted in a metal can which is carefully grounded to protect the user from shorts in the high voltage system. Additional protection is provided by enclosing the high voltage wiring in a grounded shield.

The miniature cathode ray tubes have proven easy to drive. They use electrostatic deflection and focussing. Because their deflection plates require signals on the order of only 300 volts, the transistorized deflection amplifiers are of a relatively straightforward design. Complementary-symmetry emitter followers are used to drive four small coaxial cables from the amplifier to each cathode ray tube. Deflection and intensification signals for the miniature cathode ray tubes are derived from a commercial analog line-drawing display which can draw long lines in 36 microseconds (nominal) and short lines as fast as three microseconds (nominal).

The analog line generator accepts picture information in the coordinate system of the miniature cathode ray tubes. It is given two-dimensional scope coordinates for the endpoints of each line segment to be shown. It connects these endpoints with smooth, straight lines on the two-dimensional scope face. Thus the analog line-drawing display, transistorized deflection amplifiers, miniature cathode ray tubes, and head-mounted optical system together provide the ability to present the user with any two-dimensional line drawing.

Head position sensor

The job of the head position sensor is to measure and report to the computer the position and orientation of the user's head. The head position sensor should pro-

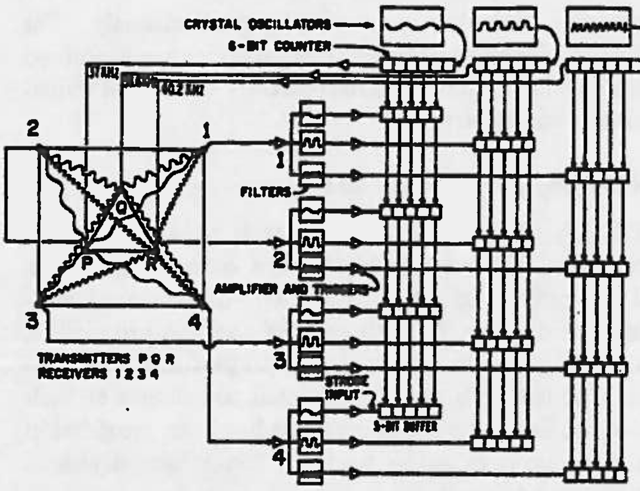

FIGURE 5—The ultrasonic head position sensor logic

40.2 kHz are attached to the head-mounted optical system. Four receivers are mounted in a square array in the ceiling. Each receiver is connected to an amplifier and three filters as shown in Figure 5, so that phase changes in sound transmitted over twelve paths can be measured. The measured phase shift for each ultrasonic path can be read by the computer as a separate five-bit number. The computer counts major changes in phase to keep track of motions of more than one wavelength.

Unlike the Lincoln Wand[12] which is a pulsed ultrasonic system, our ultransonic head position sensor is a continuous wave system. We chose to use continuous wave ultrasound rather than pulses because inexpensive narrow-band transducers are available and to avoid confusion from pulsed noise (such as typewriters produce) which had caused difficulty for the Lincoln Wand. The choice of continuous wave ultrasound, however, introduces ambiguity into the measurements. Although the ultrasonic head position sensor makes twelve measurements from which head-position information can be derived, there is a wave length ambiguity in each of the measurements. The measurements are made quite precisely within a wave, but do not tell which wave is being measured. Because the wavelength of sound at 40 kHz in air is about 1/3 of an inch, each of the twelve measurements is ambiguous at 1/3 inch intervals. Because the computer keeps track of complete changes in phase, the ambiguity in the measurements shows up as a constant error in the measured distance. This error can be thought of as the "initialization error" of the system. It is the difference between the computer's original guess of the initial path length and the true initial path length.

We believe that the initialization errors can be resolved by using the geometric redundancy inherent in making twelve measurements. We have gone to considerable effort to write programs for the ultransonic head position sensor. These programs embody several techniques to resolve the measurement ambiguities. Although we have had some encouraging results, a full report on the ultrasonic head position sensor is not yet possible.

The perspective transformation

Generating a perspective image of three dimensional information is relatively easy. Let us suppose that the information is represented in a coordinate system based on the observer's eye as shown in Figure 6. If the two-dimensional scope coordinates, X_s and Y_s, are thought of as extending from -1 to $+1$, simple geometric reasoning will show that the position at which a particular point should be displayed on the screen is related to its position in three-dimensional space by the simple relations:

$$X_s = \frac{x'}{z'} \cot\text{an}\frac{\alpha}{2}$$

$$Y_s = \frac{y'}{z'} \cot\text{an}\frac{\alpha}{2}$$

If an orthogonal projection is desired, it can be obtained by making the value of z' constant. Because the perspective (or orthogonal) projection of a straight line in three-dimensional space is a straight line, division by the z' coordinate need be performed only for the endpoints of the line. The two-dimensional analog line-

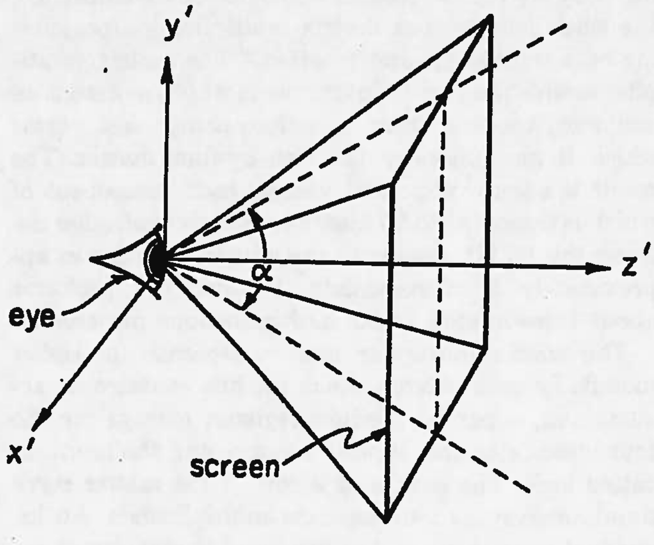

FIGURE 6—The x' y', z' coordinates system based on the observer's eye position

Results

I did some preliminary three-dimensional display experiments during late 1966 and early 1967 at the MIT Lincoln Laboratory. We had a relatively crude optical system which presented information to only one of the observer's eyes. The ultransonic head position sensor operated well enough to measure head position for a few minutes before cumulative errors were objectionable. The coordinate transformations and perspective computations were performed by software in the TX-2. The clipping operation was not provided: if any portion of a line was off the screen, the entire line disappeared.

Even with this relatively crude system, the three dimensional illusion was real. Users naturally moved to positions appropriate for the particular views they desired. For instance, the "size" of a displayed cube could be measured by noting how far the observer must move to line himself up with the left face or the right face of the cube.

Two peculiar and as yet unexplained phenomena occurred in the preliminary experiment. First, because the displayed information consisted of transparent "wire-frame" images, ambiguous interpretations were still possible. In one picture a small cube was placed above a larger one giving the appearance of a chimney on a house. From viewpoints below the roof where the "chimney" was seen from inside, some concentration was required to remember that the chimney was in fact further away than the building. Experience with physical objects insisted that if it was to be seen, the chimney must be in front.

A second peculiar phenomenon occurred during the display of the bond structure of cyclo-hexane as shown in Figure 8. Observers not familiar with the rippling hexagonal shape of this molecule misinterpreted its shape. Because their view of the object was limited to certain directions, they could not get the top view of the molecule, the view in which the hexagonal shape is most clearly presented. Observers familiar with molecular shapes, however, recognizedthe object as cyclo-hexane.

In more recent experiments with the improved optical system and vastly improved computation capability, two kinds of objects have been displayed. In one test, a "room" surrounding the user is displayed. The room is shown in Figure 9 as it would look from outside. The room has four walls marked N, S, E, and W, a ceiling marked C and a floor marked F. An observer fairly quickly accommodates to the idea of being inside the displayed room and can view whatever portion of the room he wishes by turning his head. In another test a small cube was displayed in the center of the user's operating area. The user can examine it from whatever side he desires.

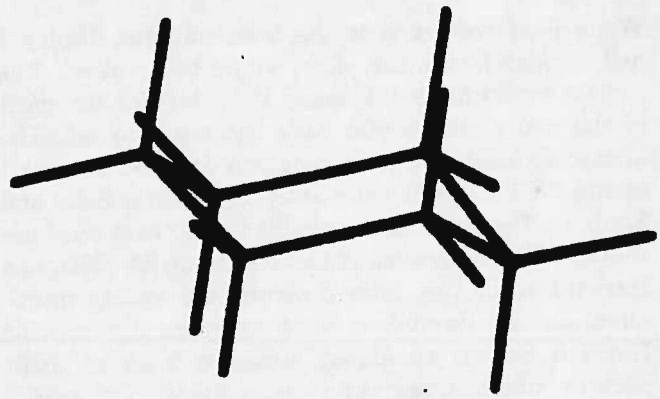

FIGURE 8—A computer-displayed perspective view of the cyclo-hexane molecule

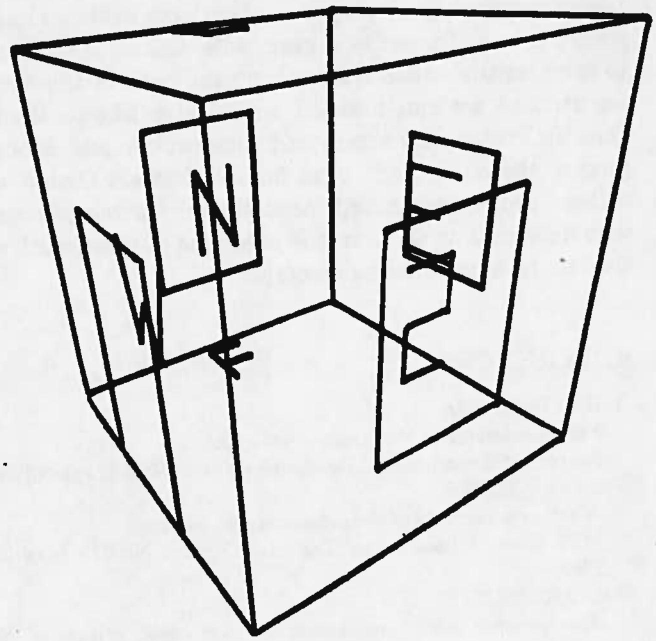

FIGURE 9—A computer-displayed perspective view of the "room" as seen from outside

The biggest surprise we have had to date is the favorable response of users to good stereo. The two-tube optical system presents independent images to each eye. A mechanical adjustment is available to accommodate to the different pupil separations of different users. Software adjustments in our test programs also permit us to adjust the virtual eye separation used for the stereo computations. With these two adjustments it is quite easy to get very good stereo presentations. Observers capable of stereo vision uniformly remark on the realism of the resulting images.