Twinkle Box—A three-dimensional computer input device

by ROBERT P. BURTON

Bell Telephone Laboratories
Holmdel, New Jersey

and

IVAN E. SUTHERLAND

Evans and Sutherland Computer Corporation
Salt Lake City, Utah

INTRODUCTION

During the past fifteen years, use of two-dimensional computer input/output devices has become commonplace. Since the earliest uses of the light pen for target identification in air defense systems it has been obvious that two-dimensional input would be interesting and useful. A large number of two-dimensional tablets and digitizers have been developed and have come into quite effective use. These devices have made use of mechanical, electrical, magnetic, optical, and acoustic phenomena. (See bibliographical references.)

More recently, the use of three-dimensional computer output devices has become prominent. It seemed obvious at first that corresponding three-dimensional computer input devices might be interesting and useful, but there has been no corresponding development of these devices. Rather, there has been a series of laboratory developments each with limited utility. A three-dimensional version of the Science Accessories Corporation acoustic tablet is the only commercially available three-dimensional input device with which we are acquainted.[9] Why has there been no prominent development of three-dimensional input devices? Aside from the obvious reason that three-dimensional graphics is used less than two-dimensional graphics, there are three reasons why three-dimensional input is not more widespread. First, early experiments with three-dimensional graphics have shown that people are not very good at drawing in space without the support of a writing surface. The years of training that grade school children go through in learning to write do not facilitate three-dimensional input. Second, the measurement of three-dimensional positions is substantially more difficult than that of two-dimensional positions. The measuring signals must travel through free space; the measuring device cannot be embedded in a surface. Finally, the coordinate conversion required to reduce the measurements actually taken to Cartesian coordinates is generally much more complex for a three-dimensional device.

This paper describes another laboratory development of limited utility.[2] We think this new device is interesting because, unlike previous devices which have measured the position of only a single point, it measures the positions of many three-dimensional points in such rapid succession that they appear to be measured simultaneously. We also feel that our novel approach to the coordinate conversion problem may be useful to others.

BACKGROUND

The earliest three-dimensional computer input device with which we are familiar is the Lincoln Wand, demonstrated by Lawrence G. Roberts in 1963.[8] Roberts' device used an ultrasonic signal and four microphones mounted at the corners of a rectangle. The path lengths from the point source of sound to each microphone were determined by the arrival times of the pulse at each microphone. As Roberts showed in his paper, the rectangular arrangement of microphones made the coordinate conversion problem fairly easy.

Concurrent with Roberts' work, Jack Raffel at Lincoln Laboratory proposed a photosensing device. Raffel's idea (never published) was to measure the ratio of the illumination falling on two photocells placed at right angles to one another. This ratio is related to the angle of arrival of the light. The position of the light could be determined from three such measurements. As will be seen later in this paper, the coordinate conversion involved would not have been particularly difficult if handled by matrix methods. However, these methods were not available to Raffel at the time.

An activity at M.I.T.[10] for measuring three-dimensional position used the critical angle of acoustic radiation into three mutually perpendicular solid rods to obtain a measurement directly in the Cartesian coordinate system defined by the rods. The idea was that the difference of arrival times at the ends of each rod locates the pulse source on a plane perpendicular to the rod and at a distance from its center proportional to the difference measured.

More recently, A. Michael Noll has developed an electromechanical three-dimensional input device in conjunction

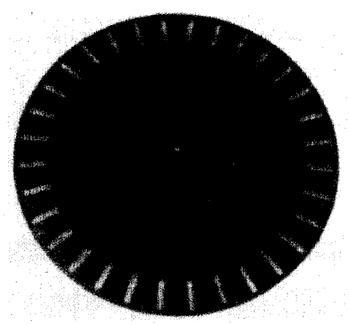

Figure 1—Disk

with a stereo display.[6] This device permits direct input in Cartesian coordinates and apparently works quite well. A group under Frederick Brooks at the University of North Carolina has been working on a combined three-dimensional input/output device. The device receives input from remotely controlled mechanical limbs, such as those used in handling radioactive materials.[3] The device can be moved by man and by the computer, serving not only for three-dimensional input, but also as a force display.

The most successful device now available is a three-dimensional adaptation of the Science Accessories Corporation acoustic tablet.[9] This device measures the time required for sound to travel from a small spark source to each of three mutually orthogonal linear microphones. The arrival times indicate the position of the spark gap. However, the Science Accessories device has several weaknesses due primarily to the slow speed with which sound travels in air. Sampling is limited to about one hundred measurements per second for a reasonable working volume. Accuracy is limited to one part

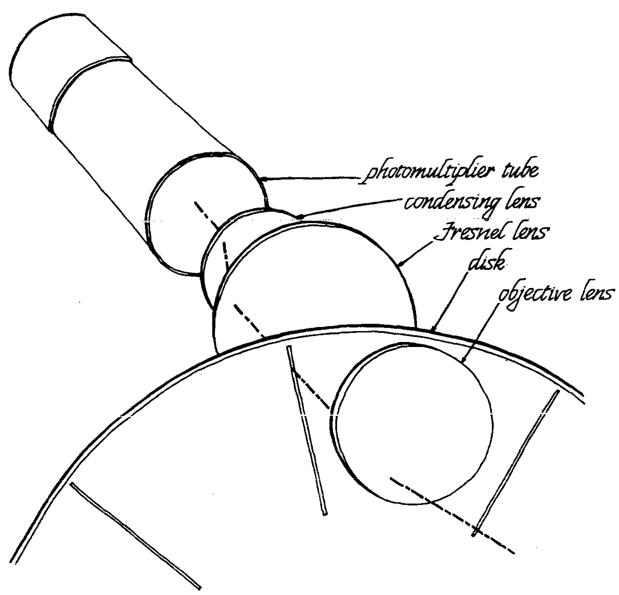

Figure 2—Optical arrangement

in five hundred due to the movement of air even in a quiet room. Finally, any object intervening between the sound source and a microphone will destroy the measurement.

In addition to some difficulties in signal-to-noise ratio, accuracy, reliability, utility, etc., each of these devices measures the position of only a single point which may be moved. Our own interest in three-dimensional input developed from use of the head-mounted display[12] for which the position and orientation of the user's head must be sensed. For this purpose, the positions in space of at least three points must be measured simultaneously. Our early efforts to make these measurements acoustically (reported in Reference 12) were never satisfactory. Instead, a mechanically coupled head-position sensor has been used. The Twinkle Box, with three lights attached to a cap, could replace the bulky, mechanical headgear. Other lights could be attached to the fingertips or body to allow the user to interact with objects viewed through the head-mounted display.

The ability to sense the positions of many points in space provides for a new kind of three-dimensional input. Rather than drawing with the point of a three-dimensional pencil, a user might make broad gestures using his fingers separately. He might grasp objects to move them, indicate sizes by gesturing with his two hands, or otherwise make use of the many three-dimensional motions with which humans (particularly Frenchmen and Italians) are said to communicate. Possible use in animation comes to mind as a result of the growing capability of computers to provide realistic perspective pictures. The ability to effectively measure real body motions with a device such as the Twinkle Box should materially aid in defining the kinds of (realistic) motions which could be imparted to the animated objects or characters. We are hopeful that the ability to measure many points in space will overcome the well-known inability of people to draw in three dimensions with a free stylus.

DETECTORS

Most people who see the Twinkle Box immediately ask why television camera technology has not been used. The

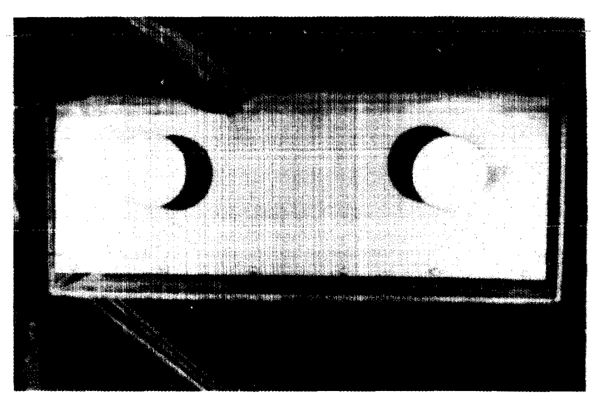

Figure 3—Detector-pair and housing

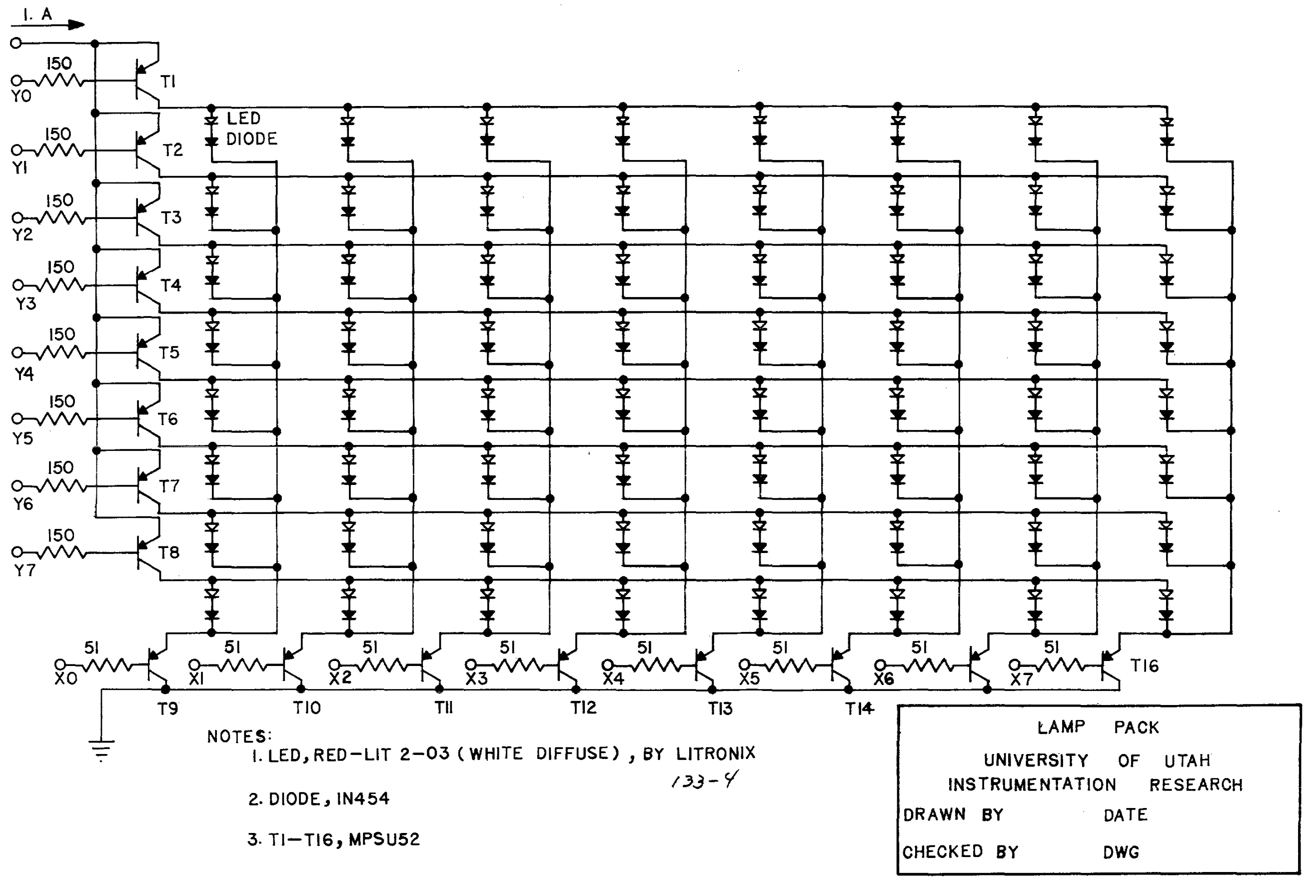

Figure 4—Diode matrix

choice of one-dimensional scanning rather than two-dimensional television type scanning is, in fact, the principal idea in the Twinkle Box design. With similar video bandwidth, hundreds of one-dimensional scans can be made in the time required for a single two-dimensional scan. The higher scanning rate makes it possible to distinguish among many light sources by turning on only one light during any one scan. Because only a single light is on at any one time, one-dimensional scans from at least three locations provide an unambiguous measure of position. To measure the positions of an equivalent number of lights using two-dimensional scanning would require a very complicated program which could match up the individual lights seen in one TV image with those seen in another image, since many lights would appear simultaneously in each image.

The Twinkle Box scanners are mechanical. Each detector-pair uses a 22-inch diameter disk with 32 radial slits cut near the edge (Figure 1). The disk rotates at 3500 rpm to provide for 1900 scans each second. Four detector-pair units are mounted in the four upper corners of a room to provide for full coverage of the room. An improved design would use some kind of electronic scanning, but mechanical scanning is sufficient to demonstrate capability.

A wide angle lens in front of the scanning disk forms a two-dimensional image of the room in the plane of the disk. When a single light is activated, this image is a single point of light which can pass through the disk only when a slit is properly positioned. Behind the disk, a Fresnel lens and a condensing lens gather light which has penetrated the slit and direct it into a photomultiplier. This optical arrangement is shown in Figure 2.

For a given orientation of the disk, the photomultiplier is sensitive to any light which lies in a plane defined by a particular slit and the center of the objective lens. As the disk rotates, the plane sweeps through the working volume of the detector. The coefficients of the plane equation are determined by the position of the slit which is determined by the time at which a pulse of light is sensed. The times at which each of three detectors sees a particular light determine three plane equations whose simultaneous solution is the position of the light.

Because the plane of sensitivity is determined by the

Figure 5—Lamp pack and user

position of a radial slit and the center of the objective lens, one might think that accurate knowledge of the disk dimensions, the focal length of the lens, the axis of rotation, and the position of the detector would be required to determine a plane equation relative to a standard reference frame. Moreover, one might think that very complicated geometric computations would be involved. As we shall see in a later section, all of the necessary geometric unknowns can be expressed in a single matrix problem. Moreover, the requisite geometric measurements can be performed at once by a simple calibration procedure using only the known positions of seven or more lights. The calibration procedure directly determines not only the relative positions and orientations of the detectors in the room, but also the effects of the focal length of the objective lenses, and the positions and orientations of these lenses relative to the associated axes of disk rotation.

Some reference time must be established if the time at which light strikes the photomultiplier is to be converted into a slit position. An auxiliary photodetector with a fixed light source determines the time at which a slit reaches a reference position during each scan. This reference assembly also provides an input for measuring variation in rotational speed.

Two detectors share each rotating disk. The two detectors are placed roughly 90° apart around the periphery of the disk so that their scanning planes are approximately at right angles to one another. The two detectors are placed near the lower left and lower right parts of a disk. With the detectors mounted just below ceiling level (Figure 3), each detector can view a large volume of the room. A pair of detectors is mounted in each upper corner of the room which measures roughly twenty feet on a side. Wide angle lenses with fields of view of approximately 90° provide complete coverage of the volume of the room.

LIGHT SOURCES

The high switching rate required to turn on each light for only one scan virtually requires that light-emitting diodes be used. Since the duty cycle of each light is relatively low, the light-emitting diodes can be severely overdriven. Unidirectional conductivity permits an 8×8 array of light-emitting diodes to be driven with just 16 drivers (Figure 4). A lamp pack with the necessary drive circuitry and jacks for eight groups of eight lights each has been assembled. The lightweight lamp pack may be worn on a user's belt (Figure 5).

There is a tradeoff between scanning rate, resolution, and the amount of light which falls on a photocell. To get a high scanning rate, the disk turns at 3500 rpm. To get high resolution, we have made the slits quite narrow, 0.3 mm in a 35 mm image. As a result, very little light actually gets from a single lamp into the photocell. Adequate sensitivity in solid state photodetectors was not available, so photomultipliers are used. Efforts to maximize lamp brightness and photomultiplier sensitivity have payed off in acceptable performance. Of course, a design using electronic rather than

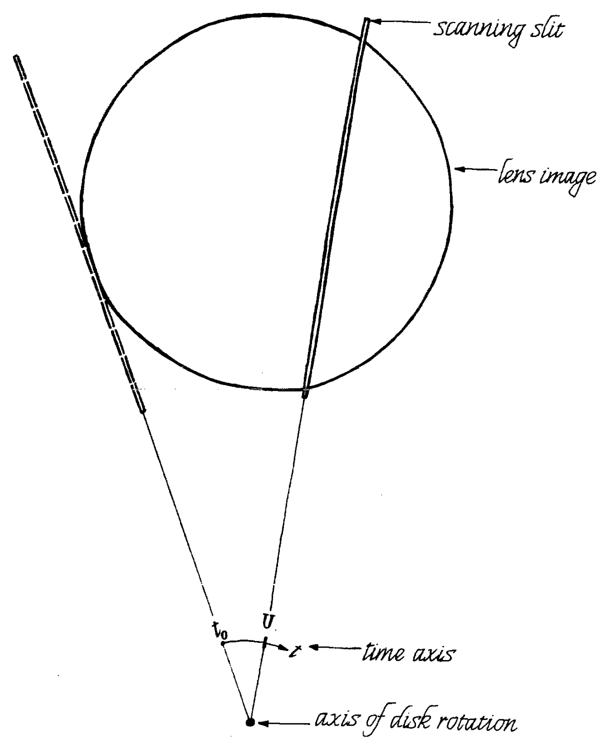

Figure 6—Perspective projection from the lens image into time space

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 \\ X_2 & Y_2 & Z_2 & 1 \\ \cdot & & & \\ \cdot & & & \\ \cdot & & & \\ X_n & Y_n & Z_n & 1 \end{bmatrix} \begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \\ \vdots & \\ t_{41} & 1 \end{bmatrix} = \begin{bmatrix} w_1 & & & & \\ & w_2 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & & w_6 \end{bmatrix} \begin{bmatrix} U_1 & 1 \\ U_2 & 1 \\ \cdot & \\ \cdot & \\ \cdot & \\ U_n & 1 \end{bmatrix}$$

Figure 7—Calculation of the transformation matrix T

or

$$\begin{bmatrix} X_1 & Y_1 & Z_1 & 1 & -X_1U_1 & -Y_1U_1 & -Z_1U_1 \\ X_2 & Y_2 & Z_2 & 1 & -X_2U_2 & -Y_2U_2 & -Z_2U_2 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ \cdot & & & & & & \\ X_n & Y_n & Z_n & 1 & -X_nU_n & -Y_nU_n & -Z_nU_n \end{bmatrix} \begin{bmatrix} t_{11} \\ t_{21} \\ t_{31} \\ t_{41} \\ t_{12} \\ t_{22} \\ t_{32} \end{bmatrix} = \begin{bmatrix} U_1 \\ U_2 \\ \cdot \\ \cdot \\ \cdot \\ U_n \end{bmatrix}$$

Figure 7(b)

or

$$\begin{bmatrix} X \end{bmatrix} \begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} U \end{bmatrix}$$

Figure 7(c)

mechanical scanning would have to face these same sensitivity problems.

COORDINATE CONVERSION

A simple matrix formulation of the computation of Cartesian coordinates from the time measurements is possible because a perspective projection is involved. Obviously the projection from the room space through the lens into the plane of the disk is a perspective projection. We can think of this image as being further projected into the space of time measurement by a projection centered at the axis of the spinning disk (Figure 6). This is again a perspective projection. In fact, the total projection from room coordinates into time space is a perspective projection. Using homogeneous coordinates it may be represented as

$$[X \ Y \ Z \ 1] \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \\ t_{41} & t_{42} & t_{43} & t_{44} \end{bmatrix} = [x \ y \ z \ w] \qquad (1)$$

and

$$U = \frac{x}{w}$$

where $[X \ Y \ Z \ 1]$ represents the room coordinates of a light, the t_{ij} are terms in a matrix related to the position, orientation, and dimensions of a detector, $[x \ y \ z \ w]$ are intermediate variables, and U is the time measurement made for that light.

Only a single measurement is made on the final perspective projection. Hence, only a single output variable U exists. The intermediate variables y and z play no part in the expression and these two columns of the matrix are irrelevant. As we learned in digitizing photographs[7] it is convenient to drop

$$\begin{bmatrix} & X^t & \end{bmatrix}\begin{bmatrix} X \end{bmatrix}\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} & X^t & \end{bmatrix}\begin{bmatrix} U \end{bmatrix}$$

or

$$\begin{bmatrix} X^tX \end{bmatrix}\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} X^t \\ U \end{bmatrix}$$

or

$$\begin{bmatrix} T \end{bmatrix} = \begin{bmatrix} X^tX \end{bmatrix}^{-1}\begin{bmatrix} X^t \\ U \end{bmatrix}$$

Figure 8—Least mean squared error fit

these two columns and rewrite the expression as:

$$[X\ Y\ Z\ 1]\begin{bmatrix} t_{11} & t_{12} \\ t_{21} & t_{22} \\ t_{31} & t_{32} \\ t_{41} & t_{42} \end{bmatrix} = w[U\ 1] \qquad (2)$$

and

$$U = \frac{x}{w}$$

which can be written

$$t_{11}X + t_{21}Y + t_{31}Z + t_{41} = U(t_{12}X + t_{22}Y + t_{32}Z + t_{42}). \qquad (2a)$$

The final expression can be thought of in three ways, depending upon what is known. If we know the position of a detector (the t_{ij}) and the position of a light $(X\ Y\ Z\ 1)$, we can use this expression to compute the time U (a useless computation). If we know the position of the light $(X\ Y\ Z\ 1)$ and the time measurement U we can compute the coefficients of one equation involving the unknown elements of the matrix T, and have the basis of a calibration procedure. Finally, if we know the position of a detector (the t_{ij}) and a time U, we can equally well compute the coefficients of a plane equation involving X, Y, and Z.

$$(t_{11} - t_{12}U)X + (t_{21} - t_{22}U)Y + (t_{31} - t_{32}U)Z + (t_{41} - t_{42}) = 0$$

or

$$aX + bY + cZ + d = 0 \qquad (2b)$$

This is the basic equation for coordinate conversion.

The calibration procedure deduces the elements of the matrix T from the times at which seven or more lights with known $X\ Y\ Z$ coordinates were sensed. The system of equations is shown in Figure 7. Since the scale factor of the t_{ij} is arbitrary, one of the t_{ij} may be specified and seven equations suffice. When more than seven reference lights are used, a least mean squared approximation is made to the resulting system of equations, as shown in Figure 8. This additional input avoids ill-conditioning in the system and reduces the effects of errors in the measurement of the positions of the reference lights. Since the values of the t_{ij} do not change, the calculation of the t_{ij} and the 7×7 matrix inversion implied in Figure 7 need be performed only once.

The coordinate conversion procedure to determine the three-dimensional position of a light source involves the simultaneous solution of the plane equations determined by each detector. As light sources move about the room each is visible to different detectors. As long as at least three detectors see a light, three plane equations can be determined and the position of the light can be deduced. If more than three detectors see a light, a least mean squared error fit can be computed.

$$\begin{bmatrix} \Sigma a_i^2 & \Sigma a_ib_i & \Sigma a_ic_i \\ \Sigma b_ia_i & \Sigma b_i^2 & \Sigma b_ic_i \\ \Sigma c_ia_i & \Sigma c_ib_i & \Sigma c_i^2 \end{bmatrix}\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \Sigma a_id_i \\ \Sigma b_id_i \\ \Sigma c_id_i \end{bmatrix} \qquad (3)$$

where the a_i, b_i, c_i, and d_i are the plane coefficients for the ith detector. For example, $a_i = t_{11} - t_{12}U$ where the t's are elements of the previously computed matrix for the ith detector, and the time U is the time measured by the ith detector. Note that the plane coefficients are simply linear combinations of the t's and U, and thus are easily found from time measurements.

Because of the arrangement of detectors and the least mean squared error fit which accommodates redundant information, no difficulties are caused by parallel or nearly parallel planes. Most of the calculation time is involved in computing the summations of Equation 3. Note that we have not incurred the difficult calculations which one might expect to be associated with the complex motion through space of several scanning planes determined by several detectors whose positions, orientations, and dimensions are arbitrary. In practice, however, even the simple calculations which are required cannot be handled by a general-purpose computer sufficiently fast to keep up with the very many simultaneous measurements of lights. Equipment appropriate to handle these computations could be built easily. At present the problem is solved by generating information no faster than it can be handled.

OPERATING EXPERIENCE

In the short time that the Twinkle Box has been in use, several demonstrations of its capability have been made. No particular technical ability has been required to use the device. Real-time sensing and conversion of a single point source of light has been shown to be quite practical. However, no practical applications have been made. Positions of multiple lights have been determined in real time at a rate of 61 points per second. Data recording for 925 light positions per second with off-line computation of Cartesian coordinates has also been accomplished. The accuracy of the system has been determined by moving lights about the room at fixed distances from one another, and measuring separation. The standard deviation of the error from zero has been determined to be 7.3 mm, due primarily to time jitter on the photomultiplier pulses.

The design has three major deficiencies resulting from the mechanical scanners. First, in order to spin the 22-inch disk at 3500 rpm, a two-horsepower motor is required. A great deal of noise, vibration, and heat is generated. With four motors running, the room becomes unpleasant to work in. Second, the high starting load presented by disk windage in a housing which does not fit tightly about the disk forced us to use induction motors. Because the disks do not rotate phase-locked, each light must be activated for the time period of two scans (one millisecond). This guarantees that the light is on at the beginning, during, and at the end of at least one complete scan by each detector. Were the disks to rotate phase-locked, we could double the data collection rate. Finally, in spite of considerable care in the production and assembly of the disks, pulse times determined by different slits for a stationary light are distributed in time. Because the pattern of distribution repeats for each revolution of the disk, we attribute it to nonconcentricity of the disk and motor-shaft, and to errors in slit placement. A correction table has been built and is used at run-time after several sequential reference pulses are received. A design allowing for continuous input by each reference sensor, to indicate which slits are scanning the images of the room, would obviate this deficiency.

There is considerable optical distortion in the wide-angle lenses chosen for the detectors.* This distortion causes inaccuracies in the measurement of positions. The magnitude of these inaccuracies is a function of the position of a light source relative to the calibration lights and relative to the optical axes of the objective lenses. Measurements near the calibration lights and near the intersection of several optical axes are quite accurate; accuracy deteriorates with increased separation. Efforts have been made to correct lens distortion. However, only a single dimension is measured by each detector. Unless corrections are made using data from more than one detector, these efforts can at most decrease the magnitude of errors by a factor of three.

An additional source of error is attributed to the fact that different detectors may see a light at different times. In this case, nonsimultaneous data are gathered. Similarly, different lights are measured at different times. Obviously, the actual time at which a detector sees a light is known, as is the time at which various lights are measured, for this is the basic measurement. Some correction might presumably be made.

* Vivitar Auto-Preset, f.1. 20 mm, F3.8 lenses are used. Linear distances are diminished up to 5.89 percent by regular distortion at the edge of a lens and up to ±1.2 percent by irregular, randomly located distortion.

POSSIBLE FUTURE DEVELOPMENTS

Henry Fuchs* has discovered that reflected light from a laser beam provides an adequate input to the Twinkle Box detectors. It is possible, therefore, to think of a device which would deflect a laser beam in two dimensions. The Twinkle Box could then sense the reflected light in one dimension. Such a device could easily measure the three-dimensional profiles of objects such as people's faces, and thus provide a new form of input. Two-dimensional deflection of a laser beam using mirror galvanometers appears to be quite practical. Since the Twinkle Box detectors are placed about the room, the object(s) to be scanned could be positioned at random.

It is possible to think in terms of reversing the Twinkle Box detectors and light sources. In such a reversed system, several one-dimensional scanning projectors would provide illumination in sequence to small photocells whose positions would be measured. Each projector would sweep the working volume with a plane of light moving in one dimension, and each photocell would report the times at which it sensed light. The mathematical computations required for coordinate conversion would be identical to those for the Twinkle Box. There are several possible advantages to such a reversed system. First, as much light as desired could be transmitted. Second, because parallel sensing of light pulses at each photocell could be provided, a much lower scanning rate could be used to measure as many positions as there are photocells. Using simpler scanners, scanning rates of 100 per second rather than 2000 would suffice. Finally, because no room image would be formed in a detector, the need for wide angle lenses with their associated distortion could be avoided. Light columnization could be accomplished by baffles and lenses. The disadvantage of the reversed system is the increased electronic complexity of providing 50 to 100 photocells each capable of reporting the time of a light pulse. Modern solid state photocells, amplifiers, and integrated circuitry appear, however, to make such a design worth serious consideration.

* University of Utah.

BIBLIOGRAPHY

1. Brenner, F. H. and M. T. Zayac, *A Multi-Color Plasma Panel Display*, Owens-Illinois, 1971.
2. Burton, R. P., *Real-Time Measurement of Multiple Three-Dimensional Positions*, University of Utah Computer Science Technical Report UTEC-CSc-72-122, June 1973. Also Technical Manual TMAN-73-01. (Supported by ARPA Contract F30602-70-C-0030.
3. Capowski, J. J., *Remote Manipulators as a Computer Input Device*, Department of Computer & Information Science, University of North Carolina (Chapel Hill), 1971. (Supported by AEC contract #AT-40-10-3817.)
4. Curry, J. E., "A Tablet Input Facility for an Interactive Graphics System," *Proceedings International Joint Conference on Artificial Intelligence*, Walker, D. E., L. M. Norton, eds. May, 1969.
5. Davis, M. R. and T. O. Ellis, "The Rand Tablet: A Man-Machine Graphical Communication Device," *FJCC 1964*, Spartan Books, Baltimore, Md.

6. Noll, A. M., "Man-Machine Tactile Communication," *SID Journal*, July/August, 1972. Also Ph.D. dissertation, Department of Electrical Engineering, Polytechnic Institute of Brooklyn, 1971.
7. Parke, F. I., *Computer Generated Animation of Faces*, University of Utah Computer Science Technical Report UTEC-CSc-72-120, June 1972.
8. Roberts, L. G., *The Lincoln Wand*, M.I.T. Lincoln Laboratory Report, Lexington, Mass., June, 1966.
9. Science Accessories Corporation, *Graf/Pen Sonic Digitizer*, Science Accessories Corporation, Southport, Conn., 1970.
10. Stoutemyer, D. R., *Systems Study and Design of a Blind Mobility Aid Simulator*, Master's thesis, Department of Mechanical Engineering, M.I.T., 1965.
11. Sutherland, I. E., "Sketchpad: A Man-Machine Graphical Communication System," *Proceedings of the Spring Joint Computer Conference*, Detroit, Mich. May 1963, and *M.I.T. Lincoln, Laboratory Technical Report No. 296*, January, 1963.
12. Sutherland, I. E., "A Head-Mounted Three-Dimensional Display," *AFIPS Conference Proceedings*, Vol. 33, pp. 757-764, December, 1968.
13. Teixeira, J. F. and R. P. Sallen, "The Sylvania Tablet: A New Approach to Graphic Data Input," *SJCC 1968*, Thompson Books, Washington, D.C.