N. Storey and J.F. Craine “Interactive Stereoscopic Computer Graphic Display Systems”, *Proc. Interact '84 pp163-168*, London, 4-7 September, 1984

# INTERACTIVE STEREOSCOPIC COMPUTER GRAPHIC DISPLAY SYSTEMS

Neil Storey, J. Ffynlo Craine
Department of Engineering,
University of Warwick,
Coventry, U.K. CV4 7AL

Conventional CRT displays give only a two dimensional representation of three dimensional objects which means that depth information can be presented only indirectly, using such techniques as hidden line removal, shading and object rotation. Improved perception of 3D images is possible by presenting to the viewer a stereoscopic pair of two-dimensional images. To be truly effective, the image presented by a stereoscopic display should change as the viewing point moves. Such a system requires a knowledge of the user's head position and orientation in order to compute the view to be presented to each eye. The paper describes an approach to a system of this type.

## 1. INTRODUCTION

The explosive growth in the use of computer aided design during recent years has brought with it the need to be able to display many kinds of three dimensional objects as realistically as possible. Unfortunately, no satisfactory technique exists yet for presenting truly three dimensional images of objects, but a wide range of optical and psychological phenomena have been exploited to give the illusion of three dimensions using currently available technology. For two dimensional displays these include such depth cueing techniques as hidden line removal, shading and object rotation. However, these techniques ultimately present only an illusion of depth. In time true 30 displays based upon holographic techniques may overcome these problems [2][18]. However, the generation of holograms from computer-generated data in real time presents considerable technical problems [17], and holographic displays seem likely to remain prohibitively expensive for the majority of engineering applications for some time.

## 2. DEPTH PERCEPTION

When considering the design of a display system for three dimensional objects, various factors associated with vision and the perception of depth must be considered. There are many "visual depth cues", and in analysing any scene it is likely that a number of cues would be used. However, the more important visual depth cues probably include the following:

(1) Motion parallax. When an observer moves his or her head, objects which are closer appear to move more rapidly than those more distant. Similarly, if two objects are moving with the same speed, the object closer to the viewer will appear to move faster than the one further away. Thus rotating an object on a conventional display provides the viewer with more information about the three dimensional shape of the object.

(2) Kinetic and static interposition. The way in which one object may partially obscure a view of a more distant object provides information as to the relative distances of the two objects , which is why hidden line removal is important for the realistic display of three dimensional objects. This effect is enhanced when one or both objects are moving.

(3) Kinetic vector effect. When the viewer and scene move steadily with respect to one another, the continuous manner in which the view changes provides information about the scene.

(4) Linear Perspective. Parallel lines appear to converge at some distant point, providing another important clue as to distance.

(5) Stereopsis. When observing a real object, each eye is presented with a slightly different view, because of their physical separation. As the line of sight of each eye converges at the observed object, its range may be determined by triangulation.

(6) Size constancy. The brain has prior knowledge of the size of many objects, and it can use this knowledge to estimate the range of an object from its apparent size.

(7) Accommodation. The degree of focussing necessary in the eye to produce a clear retinal image provides a cue to an object's distance.

(8) Texture gradation and shading. There is a gradation of texture from rough in the foreground to smooth in the distance. Similarly, shading depends upon the intensity of illumination. This provides information about the angle presented by a surface to the light falling upon it, and may also provide more direct information about distance if the intensity of illumination varies with position.

(9) Aerial perspective. The clarity of more distant objects may be reduced by environmental effects such as haze and mist.

The relative importance of each effect depends upon the particular scene and possibly upon the individual observer.

## 3. EXPLOITING DEPTH CUES IN COMPUTER DISPLAYS

Some of these depth cues are used by artists , who have long been faced with the problem of portraying three dimensional views in two dimensions [3] [4]. We are accustomed to interpreting two dimensional illustrations of three dimensional objects. In doing this we make use of size constancy, perspective, texture gradation, shading and aerial perspective. The generation of an image by a computer may require a large amount of computation.

Simple display systems make use of only the most elementary depth cueing methods such as perspective and hidden line removal. More sophisticated systems add to these the use of shading and aerial perspective to improve the illusion of depth, but at the expense of a considerable increase in computational complexity, especially if curved edges are to be reproduced. In practice the limited computational power available often means that only static images can be drawn. Even the very best static image could be compared only with a painting or photograph, in that many potentially valuable depth cues are missing. When motion can be displayed, kinetic effects such as motion parallax and kinetic interposition can be exploited to give a more realistic illusion of three dimensionality. However, kinetic effects should arise not only from the motion of the object being viewed, but from the motion of the observer, and only the former is under the control of the display system. Thus, with a conventional display, if the viewer moves his or her head, he or she will simply have a different view of the same two dimensional image. This is comparable with the use of film and television to portray three dimensional scenes, and the very best dynamic computer graphic techniques could be expected to give a display comparable to a high quality television picture. However, depth cues such as those of stereopsis, accommodation and motion parallax due to head motion, are still not exploited.

In order to produce depth cues due to accommodation, a display must produce an image in which rays of light appear to come from the surface of the three-dimensional object being displayed, rather than from a flat surface. This can be achieved using holographic techniques, but these currently. p. resent considerable technical difficulties. Some existing methods for producing three dimensional images do provide depth cues due to accommodation. These include the use of moving mirrors [5] [19] [20] [21] [22]. However these systems generally create only a limited perception of depth and cannot produce correct hidden-line removal. Binocular vision provides us with important information about an object, and yet th is information is completely lo st with a two dimensional display. A stereoscopic pair of images can be generated relatively easily by extension of existing two-dimensional techniques without recourse to more exotic technology. In most cases, stereopsis provides a more important cue than accommodation.

Depth cues due to motion of the head are important when viewing real objects. As the viewpoint changes, different parts of the object can be seen and the perspective changes. This effect is particularly important when the object is close to the observer. However, although this effect is obtainable with holographic techniques, it is not produced by conventional computer display systems.

## 4. STEREOSCOPIC DISPLAYS

It has been known for over a century that perception of 3D images is possible by presenting to the viewer a stereoscopic pair of two-dimensional images. This technique is both simpler and cheaper to implement than a holographic system, and the use of stereoscopic displays can greatly enhance the interpretation and assimilation of three dimensional material.

Stereoscopic computer graphics are potentially of immense value in adding realism to images of three dimensional objects and allowing rapid and accurate visualisation of workpieces when using computer aided design. Many techniques have been developed for the presentation of stereoscopic images for photographic and television applications [3] [13] [16]. This work has generally attempted to produce non-localised images that can be seen from a range of positions and hence simultaneously by a number of viewers. Such systems have usually employed head mounted apparatus to provide the two eyes with different views of the scene being displayed. The techniques used include the use of filters to separate images of different colours (typically red and green) [10]; cross polarised filters to separate images of different planes of polarisation; and "eye-switching" spectacles which blank out the view from each eye in turn to separate images which are multiplexed in time [7] [8] [14].

Stereoscopic displays have been produced which do not require any head mounted apparatus. Some such systems work by generating two or more localised images using an optical arrangement and restricting head position so that the eyes are positioned to see an appropriate pair of images [9]. Other systems simply present both images to both eyes and rely on the "pattern recognition" abilities of the viewer to interpret the information correctly. These approaches both have associated problems: the first restricts the viewing position to correspond with that of the images, and the second is difficult to interpret for inexperienced users. It is for these reasons that the majority of applications of stereoscopic displays have used non-localised images and some form of head mounted spectacles.

Most stereoscopic displays are specifically designed so that the same view is seen from any position. However, when a person moves their head they expect, subconsciously, that their field of view will change accordingly. The fact that the stereoscopic image does not change as the viewing point is changed results in a very disturbing effect. Because "distant" parts of the image show the same angular motion as those "closer" to the viewer, the image appears to rotate as the head moves. This effect is most pronounced with those parts of the stereoscopic image which are "closest to" and "furthest from" the viewer.

A second effect is that a stereoscopic image appears to have the correct orientation and perspective only when viewed from the point for which the stereoscopic pair was calculated. When seen from any other point it appears to have the wrong orientation and to be distorted. This is a serious shortcoming when the display is being used for applications such as computer aided design where a correct perspective may be vital. Either the image must be adjusted as the viewing point changes or the viewer's head must be constrained from moving. The la tt er alternative would not appear to be particularly attractive.

To be truly effective a stereoscopic display should be position dependent since the image should change as the viewing point moves. This is difficult to achieve if the image is to be viewed by more than one person at the same time and applications such as stereoscopic film and broadcast television must generally be content with a position-independent display. However, when the image is part of a man-machine interface this restriction to a single viewer is not a serious shortcoming. Moreover, because the material to be displayed is computer generated, it is possible to calculate the appropriate views for each eye, provided that their position is known. This produces an image which appears to be stable in space, and in which distances and perspectives are correct. It also allows the viewer to move his or her head to obtain different views of an object within the limitation of the field of view of the display [6]. Clearly, such a system requires a knowledge of the user's head position and orientation in order to compute the appropriate view to be presented to each eye.

## 5. HEAD POSITION DETERMINATION

Over the years many techniques for measuring head position have been used in a number of applications. These range from rather inelegant mechanical arrangements using apparatus connected to the head by levers or wires, to highly sophisticated non-contact methods.

Most techniques have used some active or passive apparatus mounted on the head. Typical of the passive approach is the use of a stationary optical transmitter and receiver with reflectors mounted on the head, often on some form of spectacles. This technique has two main variants: either an omni-directional transmitter is used with a receiver which scans its field of view to locate the reflected signal, or the receiver is omni-directional and a scanning transmitter is used.

Examples of the use of active head-mounted apparatus are the use of light emitting diodes (LEDs) mounted on spectacles with a fixed detector or camera which can locate their position; and the complementary arrangement of head-mounted photodetectors and one or more fixed scanning light sources. These techniques have been successfully incorporated in to many systems, including military applications such as the use of head position for targeting of both airborne and ground based weapons and navigational equipment. These applications have also led to the development of the "space synchro" which uses a nutating magnetic field and a head-mounted detector [23]. The phase and direction of the magnetic field can be used to calculate the position and orientation of the detector, and hence the head. However, this technique is expensive.

Ideally, a head-position measuring system for use with a stereoscopic display should require no head-mounted equipment. This is a practical possibility using, for example, a television camera to observe the operator, and a pattern recognition system to identify the position of the eyes. .Two cameras, or a single camera with a split field of view, would allow range to be measured in addition to angular position, or this could be measured directly by optical or other means.

## 6. AN INTERACTIVE STEREOSCOPIC SYSTEM

A knowledge of head position is also useful in the display of images [6] [15]. Since the location of the eyes is known the stereoscopic images need only be local, and they can be made to "follow" the eyes using a "steerable" display which generates the calculated view for each eye in the appropriate position. This produces a stereoscopic display system which allows the user to move his head to obtain kinetic depth cues without creating distortions of the object being viewed.

An interactive stereoscopic display system would present the viewer with virtually all the visual cues listed in section 2 with the exception of accommodation. An interactive stereoscopic image of this type would offer the computer aided designer a three dimensional "drawing board". The system used for head position determination could be developed to allow the simultaneous determination of the position of a pointer which could be used as a 3D "light pen". In this way the viewer could interact with the image to make modifications or to move it within the field of view.

## 7. ARCHITECTURES

Hitherto, vector scan displays have been used for computer graphic display systems, but nowadays they are losing favour to raster scan devices [1], This is largely because the vector scan display is suited only to drawing lines (hence its name) and cannot perform functions such as polygon filling which can be done with raster displays. Consequently the vector scan technique is not particularly suited to displaying solid objects realistically.

However, the vector scan display is generally superior to its raster scan counterpart in terms of resolution. A line drawn on a raster scan display is formed from points in a fixed matrix, unlike a line drawn on a vector scan display in which only the end-points are constrained in such a way. The effects of finite resolution within any display will limit the degree of realism which is attainable. In the case of a stereoscopic display this is complicated by the fact that depth perception is achieved by comparing the positions of edges, so that quantisation of edges will affect the apparent position of the object.

A very large amount of computation is required to convert information from a description of a scene into a form suitable for connection to a television monitor [12]. Simple systems cope with this problem by means of a bit plane in which each pixel of the image is represented by a memory word. However, the amount of computation required means that a relatively long time - seconds or even minutes - is required to build up the image in memory. This time can be greatly reduced by using special LSI display controller chips, but the amount of time taken to form an image of any but the simplest kind often precludes the generation of moving pictures.

Even if a circuit were capable of calculating the state of all the pixels in the time between successive frames of a TV picture, the bandwidth of the bit plane presents a "bottleneck". Currently available dynamic RAMs have access times in the order of 150ns, to which must be added the execution time of the circuit modifying the state of each pixel. This might be expected to raise the time per pixel to the order of 200ns or more. The time available to calculate a new frame is 16.7 or 20 ms, depending upon the frame refresh rate (60 or 50Hz) of the display, which means that the maximum number of pixels which can be handled by a single process is at most about 100000. This corresponds to a resolution in the display of approximately 300 pixels in both the X and Y directions. Modern graphic displays require appreciable higher resolution, but the total number of pixels in the display rises as the square of its linear resolution, and as the image positions must be updated during each frame as the head moves, no more time is allowable for each frame. Thus a 4096x4096 pixel display would require more than 16 million words of memory with an access time in the order of one nanosecond.

The high processing speed required to produce a steerable display precludes the use of conventional bit-plane techniques in the video generator and therefore the video signals must be produced directly using very high speed logic circuitry. The simple analysis of the last paragraph is, of course, very approximate. For example, it excludes the time taken to transform the image from its descriptive form within the display computer to the display output, which requires a considerable amount of arithmetic in order to "project" a three dimensional object to a stereoscopic pair of two-dimensional images. Clearly the task is too great for a single processor using the technology available. However, the low cost of microprocessor technology allows several processors to be included within a system at reasonable cost. There are many ways in which a number of processors could be combined to handle the computation required in this application. A set of special purpose processors could operate in parallel, each of them being responsible for a separate part of the displayed image. This method is quite attractive for high resolution displays, because if each processor has independent access to its own memory it is not limited by the bandwidth "bottleneck" already referred to.

Another way to divide the computation involved is to construct a "pipeline" of processors, in which the output from one process is connected to the input of the next. In this way data flows through the pipeline on its way to the screen. Different types of processor can be used, the choice depending upon the function which it is to perform. Mathematically intensive operations such as transformation can be carried out by a 16-bit general purpose processor or by a high speed signal processing circuit. Some of the signal processors currently available are capable of performing a multiplication in well under one microsecond so that a large number of points can be transformed in the time available. The later stages in the pipeline require a large amount of information to be processed within the time taken to display one line of the picture, that is, about 64 microseconds. In this case, because of the high speeds required, special purpose state machines designed using logic circuits are necessary. Operating speeds of some tens of megahertz are needed for this section, a daunting but by no means impossible task.

It is not difficult to identify the main processes which must be performed in this pipeline architecture.

(1) The three dimensional scene must be projected to give the two views as seen by the two eyes. This requires input from the head position sensor and from the internal scene description. The result is a set of "patches" with attributes such as colour, shape, position in the display, and distance from the observer. This last attribute is important because it allows the effects of interposition to be displayed.

(2) Next, the points at which this patch overlaps each line of the display are calculated. These points appear in the order in which they are calculated.

(3) The points are sorted in to the order in which they will appear on the line. This requires the use of special hardware; ideally a content-addressable memory would provide a fast parallel method of sorting, but very high speed serial methods can give sufficiently rapid operation.

(4) The attribute of the surface closest to the viewer is then determined for each pixel along the line, making use of the distance attribute referred to in (1) above.

(5) This attribute must then be converted in to a colour, and perhaps an antialiasing filter used to minimise the visual impact of the quantisation caused by the display raster.

All except the first process must be carried out in two separate channels, one for each eye. Thus in addition to having a pipelined architecture, the design must also have a limited degree of parallelism. The resulting system will have about ten processors, which not long ago would have represented a complicated design. However, the availability of LSI logic circuits makes such a system feasible even for displaying quite complicated scenes [11]. The use of special purpose logic circuits and high speed logic such as ECL would allow a considerable increase in the complexity of scene available with th is technique.

## 8. CONCLUSIONS

A system for the display of animated stereoscopic images has been proposed in which the information presented to the viewer is changed as the viewpoint is altered. Work on such a system is currently under way, funded by a grant from the Science and Engineering Research Council. While the implementation of a system of this type presents considerable technical problems, it is believed that the improved visual depth cueing achievable with such a system will justify the effort required to overcome these problems.

## 9. ACKNOWLEDGEMENT

The authors wish to thank Dr. E.L. Hines and Mr. I. Sexton for their assistance in the preparation of this paper.

## 10. REFERENCES

[l] Blinn J.F., "Raster Graphics", Computer Graphics, ed Booth K.S. .IEEE 1979

[2] Sexton S.A., "Survey of Holographic Stereograms", SPIE Vol. 367 1982

[3] Nakayama K., "Geometric & physiological aspects of depth perception", SPIE Vol. 120 1977

[4] Tyler C.W., "Spatial limitations of human stereoscopic vision", SPIE Vol. 120 1977

[5] Fuchs H., "Design of Image editing with a space filling three dimensional display based on a standard raster graphics system", SPIE Vol. 367 1982

[6] Sutherland I.E., "A head mounted three dimensional display", AFIPS FJCC 1968

[7] Roese J.A. & Khalafalla A.S., "Stereoscopic viewing with PLZT ceramics", Fe rr oe le ct ri cs 1976

[8] Roese J.A. & McCleary L., "Stereoscopic computer graphics for simulation & modelling", proc. Siggraph 1979

[9] Butterfield J.F., "Autostereoscopic displays using a stereo-pair of images", SPIE Vol 162 Visual simulation and image realism, 1978

[l0] Smith C., "The secrets of television in depth", New Scientist, 21 Jan. 1982

(11] Fussel D. and Rathbone B.D. "A VLSI oriented architecture for real time raster display of shaded polygons", Graphics Interface 1982

[12] Black S.R., "Digital processing of 3-D data to generate interactive real-time dynamic pictures", SPIE Vol 120 Three-Dimensional imaging. 1977

[13] Lane B., "Stereoscopic displays", SPIE Vol 367 Processing and display of Three-dimensional data, 1982

(14] Robinson MI and Sood S.C., "Real-time depth measurement in a stereoscopic television display", SPIE Vol 367 Processing and display of Three-Dimensional data, 1982

[15] Fisher S., "Viewpoint dependent imaging: an interactive stereoscopic display", SPIE Vol 367 Processing and display of Three-Dimensional data, 1982

[16] Butterfield J.F., "Very high resolution stereoscopic television" , SPIE Vol 120 Three-Dimensional imaging, 1977

[17] Peximutter R.J., Goodman J.W. and Macovski A., "Digital holographic display of medical CT images1', SPIE Vol 367 Processing and display of Three-Dimensional data, 1982

[18] Huff L. and Fusek R.J., "Application of holographic stereograms to threedimensional data display", SPIE Vol 367 Processing and display of Three-Dimensional data, 1982

[19] Sher L.D., "3-D or not 3-D?', Proc. 15th Int , Conf. on systems sciences, Vol 2, 1982

[20] Mark H. and Hull F., "Three-Dimensional viewing of tomographic data-The Tomax system", SPIE Vol 120 Three-Dimensional imaging, 1977

[21] Rhodes M.L., Stover H.S. and Glenn W.V., jr., "True three-dimensional (3-D) display of computer data", Medical applications, SPIE Vol 120 Three-Dimensional imaging, 1977

[22] Simon W.and Walters T., "A spinning Mirror auto-stereoscopic display", SPIE Vol 120 Three-Dimensional imaging. 1477