POLYGON SHADOW GENERATION

by

Peter Atherton, Kevin Weiler and Donald Greenberg

Program of Computer Graphics

Cornell University

ABSTRACT

A general purpose method for generating shadows using a polygonal coordinate data base is presented. The method is based on an object space polygon clipping hidden surface removal algorithm. Output from the program is in the same three-dimensional polygon format as the input. Thus, a shadowed data environment may be easily created and viewed from any observer position with no additional depth sorting time required for the hidden surface removal process. Shadows can also be cast by more than one light source. Since the shadows are generated in object space, the results can be used for both visual display and numerical analysis.

COMPUTING REVIEWS CLASSIFICATION: 3.2, 4.9, 4.40, 4.41

KEYWORDS: Shadow Generation, Hidden Surface Removal, Polygon Clipping, Graphics

I. INTRODUCTION

A shadow is the darkness cast by an object intercepting light. It falls from the side opposite the source of light. Theoretically, when the observer's position is coincident with that of the light source, no shadows are visible. Shadows become visible when the viewer's position moves away from the source of illumination.

The addition of shadows to a perspective image vastly improves the depth perception of the display. Furthermore, shadows provide valuable positional information and improve the ability of the observer to comprehend complex spatial environments. However, computation times and algorithmic complexity for shadow generation have prevented many implementations. The shadow creation method presented is a natural extension of an object space hidden surface removal algorithm which uses polygon area sorting and is described in the third section. A major advantage of this method is that both the input and output are in the form of a three-dimensional polygon data structure. This characteristic means that the shadow definitions can be used for the purposes of both display and analysis.

II. SURVEY OF EXISTING ALGORITHMS

Several classes of algorithms for shadow generation have been previously presented. Each of these approaches has inherent limitations which may restrict their application and use. The raster scan method for shadow image creation was first implemented in 1970 by Kelley and Bouknight[4,5,7] although a similar procedure for line drawing images has been presented by Appel.[1]

Oriented to a raster type display scope, the Kelley and Bouknight method scans an object row by row to determine visibility. Each time a polygon boundary is crossed, a depth sort is made to determine which polygonal surface is nearest the observer (Figure 1). Since the color of a polygon does not change across its surface, the only display information necessary is the location of the "key squares," those raster units in each row where a color change takes place.

Shadows may be added to an image simply by running two concurrent scanning operations, one to determine visible surfaces and one in image space to determine shadow existence. Before scanning, a list is created for each polygon linking it to any other polygon that might cast a shadow upon it.

In 1973 Nishita and Kakamae presented a method for shadow generation based on a convex polyhedron clipping algorithm. This program maintained some of the benefits of the raster scan display method while improving on the accuracy and versatility of the shadow definitions.

The data input base consists of convex polyhedra, each of which may be composed of several convex polygons. Hidden surface removal from any chosen point of view is accomplished by determining the silhouette contours of each polyhedron and using them to define its clipping border. Objects which lay behind a selected polyhedron are clipped to the window defined by the polyhedron's outside boundaries.

The generation of shadowed images by this polyhedron clipping method is accomplished in two basic

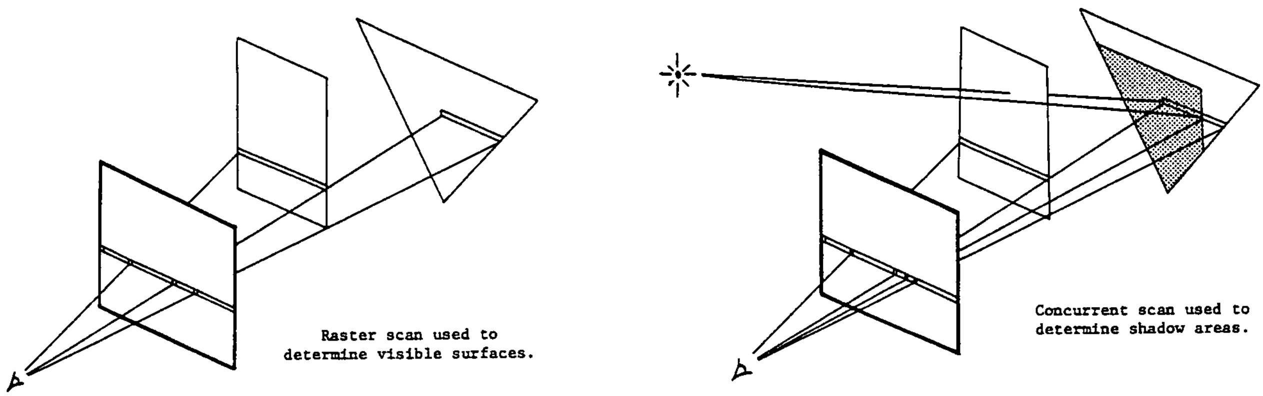


Figure 1. Concurrent scanning method of shadow display.

steps (Figure 2). In the first step, a view is taken in the direction of the infinite light source. Using the polyhedron clipper, all the hidden surfaces, which are surfaces that are in shadow, are found. The entire scene is then transformed to a selected view point, and all hidden surfaces are removed by a raster scan method similar to that used by Bouknight and Kelley.

A third algorithm for generating shadows is based on the concept of computing the surface defining the volume of space swept out by the shadow of an object, its umbra. The umbra surfaces are then added to the data and treated as invisible surfaces which, when pierced, cause a transition into or out of an object shadow. This shadow volume approach was presented by F. Crow in 1977.[6]

For any polyhedron, the shadow volume can be completely described for a given light source position (Figure 3). The contour edges of the original object, as seen from the light source, are first computed. Then all planes defined by the light source and the contour edges constitute the bounding surface of the shadow volume. The "near" surface of the shadow volume is defined by the silhouette edges of the object casting the shadow. The "far" surface is at an infinite distance. This volume is then clipped by the frustrum of vision (or viewbox) and added to the environment data base. Any hidden surface algorithm can then be used to create the display. The shadow data is treated in the same manner as the original data except that it is invisible. In the depth order calculations, any plane behind a front facing shadow surface is in shadow. The method can be coupled with several hidden surface algorithms and has the capability of effectively creating shadow volumes when the illuminating light source is placed within the original environment.

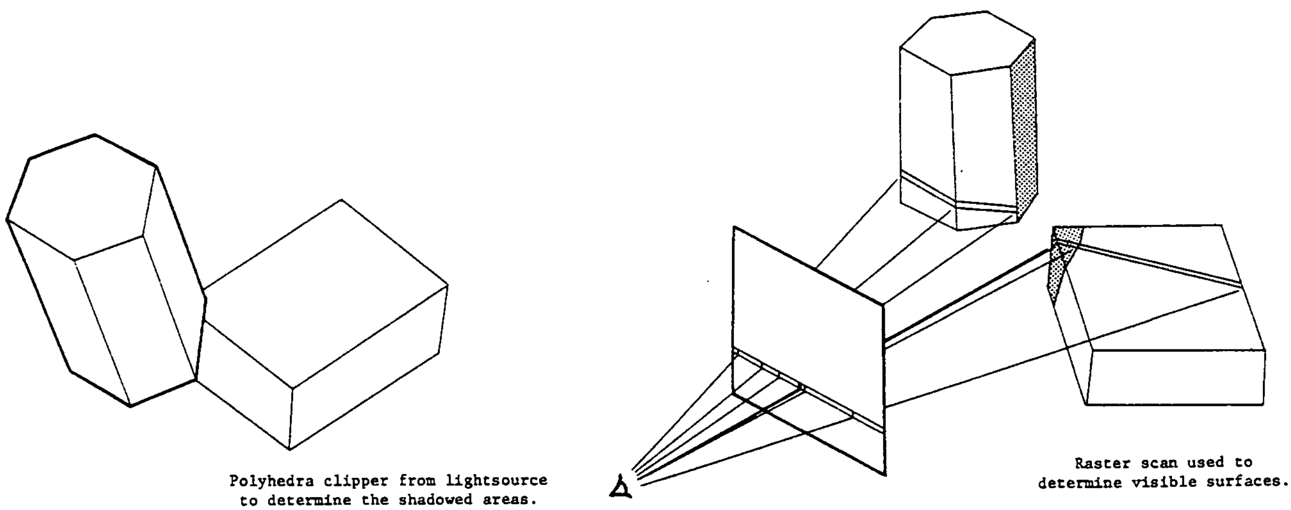


Figure 2. Polyhedra clipper and raster scan method of shadow display.

There are several restrictions with the shadow image creation programs described. The major limitation of the double scanning method and the shadow volume method is that the shadow surfaces are not defined in object space. This precludes the use of the created shadows for accurate computational purposes. Furthermore, scan-line algorithms based on raster display devices determine their depth priority in image space. This limits output portability (e.g., vector displays) and will become less efficient as the display resolution increases. Additional problems, unique to the individual algorithms, are the imposed limitations on the environment description, the potentially large increases in the environment data base, the required maintenance of non-visible polygons as possible shadow casting elements, and the necessity for recalculating shadows for each image.

To overcome these difficulties, a different approach to shadow image generation based on our hidden surface removal method using polygon area sorting has been developed.[2,10,11] This can be accomplished with relative ease since the polygon form of the output is the same as the polygon form of the input.

III. POLYGON AREA SORTING HIDDEN SURFACE REMOVAL ALGORITHM

A program to remove hidden surfaces by polygon area sorting has been developed at Cornell's Laboratory of Computer Graphics.[10,11] The basic concept of a polygon sorting hidden surface removal algorithm is that all surfaces that lay behind each unique polygonal area and within its borders are removed. The algorithm proceeds from front to back across the transformed object space, producing portions of the final image along the way and temporarily reversing direction only when an initial depth sort error is detected. Output from the algorithm never overlaps on the vertical image plane since each visible area has had all polygons behind it removed. This polygonal area may itself be subdivided recursively if there is an error in the initial depth sort.

The hidden surface removal algorithm involves four basic steps:

1) a preliminary rough depth sort
2) a two-dimensional comparison of the currently most forward polygon, or template, to the remaining polygons
3) removal of polygons that exist behind the template and within its borders
4) a recursive comparison when an error in the preliminary depth sort has occurred.

At the heart of the hidden surface removal process is a polygon clipper. This algorithm considers two polygons at a time, a template or clipping polygon and a subject polygon. The two polygons are compared and the surfaces of the subject polygon existing within the borders of the clipping polygon are designated. Even though the polygon clipper works essentially in two dimensions, all depth information is accurately preserved maintaining the precise three-dimensionality of the polygons.

The polygon clipper is capable of clipping a concave subject polygon with holes to the borders of a concave clipping polygon with holes. This generality is necessary since even when a scene is restricted to convex polygons, a clipping sequence could quickly yield concave areas and holes. Surface details such as texture or color differences can be described as polygons within the boundaries of a parent polygon. These surface details will have a minimal effect on the hidden surface removal process.

IV. POLYGON SHADOW ALGORITHM

The procedure for creating an image containing shadows consists of two major parts. The first is the creation of the shadow descriptions as dictated by the particular object orientations and light source position. The second is the determination of visible surfaces with their associated shadow descriptions and is dependent upon the observer's position.

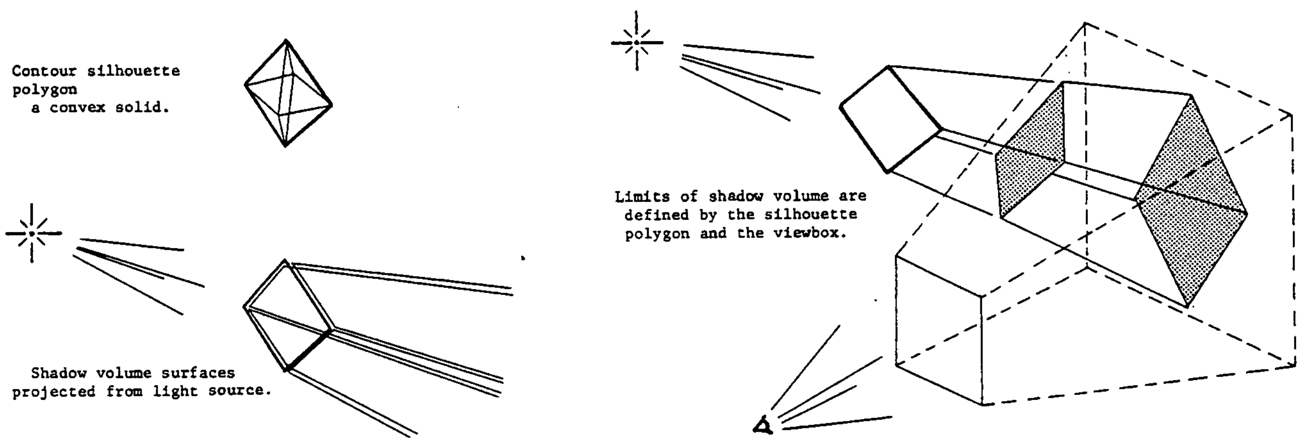


Figure 3. Shadow volume method of shadow display.

By using the general purpose polygon clipping hidden surface removal algorithm previously described, the process of generating shadowed images can be made relatively simple (Figure 4). Shadow descriptions are found by viewing the environment from the position of the light source. A hidden surface removed view from the light source position will delineate the illuminated polygons which are those areas not in shadow. Once defined, these illuminated polygons are added to the original environment and treated as surface details on their original source polygons. This general method is suitable for both point light sources and parallel light sources.

This approach has several distinct advantages. First, since the polygonal output of the hidden surface removal routines is the same as the input, the same logic can be used for the shadow generation and the image display. Second, by maintaining the three-dimensional shadow polygon output, it is possible to compute shadow areas and thus their effect on such phenomena as energy utilization. Third, by adding the shadows to the data base in the form of details attached to "parent" planes, the computational time for the hidden surface removal sorting process does not increase. Fourth, many views can be generated requiring only one original shadow generation cycle. Lastly, shadow views with multiple light sources require only a single pass through the hidden surface removal program from the viewpoint of each light source. At present, the general polygon shadow generation procedure is only limited by the requirement of a polygonal planar data base. It has proven to be flexible, device independent, and has run efficiently on a large variety of environments.

For each display frame that is to be produced, there is a set of transformation matrices which are used to manipulate the environment coordinate data. These transformation matrices are of two types, <u>view</u> matrices and <u>shadow</u> matrices. The view matrices transform the environment to any selected view. There are two shadow matrices which are devoted to the creation of a shadow data base consisting of the original polygonal coordinate definitions and their associated lighted detail polygons (Figures 4 and 5).

The first of these shadow matrices is used to transform the entire object environment to the viewpoint of the light source. A copy of the transformed environment is made for later use. Hidden surfaces are removed from the object environment leaving only the illuminated polygons. The second shadow matrix is then used to transform the entire copy of the object environment to any environment orientation including the original orientation. The lighted polygons are also transformed by the second shadow matrix and then added to copied polygonal data as lighted details to derive a shadowed coordinate data file. Once the shadowed data file is created, only one view matrix is needed to transform it to any desirable viewing position.

V. FITTING THE ENVIRONMENT INTO THE VIEWBOX

In performing the computations for the display of shadows, it is important that the object be entirely contained by the frustrum of vision emanating from the light source. Areas of the object that exist outside of the viewing area will be clipped and removed, and thus falsely interpreted to be in shadow. Therefore, the entire object must be within the boundaries of the viewing area.

By performing shadow calculations in object space with the polygon clipping method of hidden surface removal, the precision may be extended to the machine limits, rather than the display limits. If the coordinate values are stored in integer format, the maximum accuracy of the shadow calculations can be obtained when the following three criteria are met:

1) The boundaries of the viewbox of the frustrum of vision are set to correspond to the maximum machine limits (e.g., for a 16 bit computer, this corresponds to +32,767).
2) The object environment is centered within the viewbox.
3) The object environment is then scaled as large as possible to fit within the viewbox.

To accomplish this, the extreme three-dimensional coordinates of the original object are used to form the minimum rectangular solid containing the entire environment. The centroid of this volume is then centered in the viewbox and scaled as large as possible with the constraint that all portions of the bounding volume remain within the viewbox window.

VI. DISPLAY OF SHADOWED IMAGES

Since the three-dimensional polygon coordinate data is maintained to the limits of machine precision, images created by the hidden surface removal system can be displayed accurately on many different peripheral devices. The two basic types of displays used are hidden line removed vector displays and hidden surface removed halftone displays (Figure 4).

The vector displays are only concerned with drawing the lines or borders of each polygon and are inherently faster than the halftone displays. Furthermore, the display is more accurate due to the available resolution of the standard vector displays. Details may be visualized easily, but the depth perception is not nearly as effective as with the halftone displays.

For color raster displays, all visible surfaces of the environment must be rendered. This is achieved with the aid of a set of software routines which can render an arbitrary concave polygon with holes with a selected color. Colors and shades can be interactively selected or automatically computed for each polygon surface. For black and white images, the shade of gray selected for a particular polygon is dependent upon the angle between a ray extending from the light source to the polygon and the normal of the polygon. Strictly speaking, shadowed surfaces would be rendered black. To aid in image visualization, the shadowed surfaces utilize a darker gray range of the gray scale than that used for the lighted surfaces. To produce a color image, the same type of intensity scale is applied to the particular ratio of basic hues (red, green

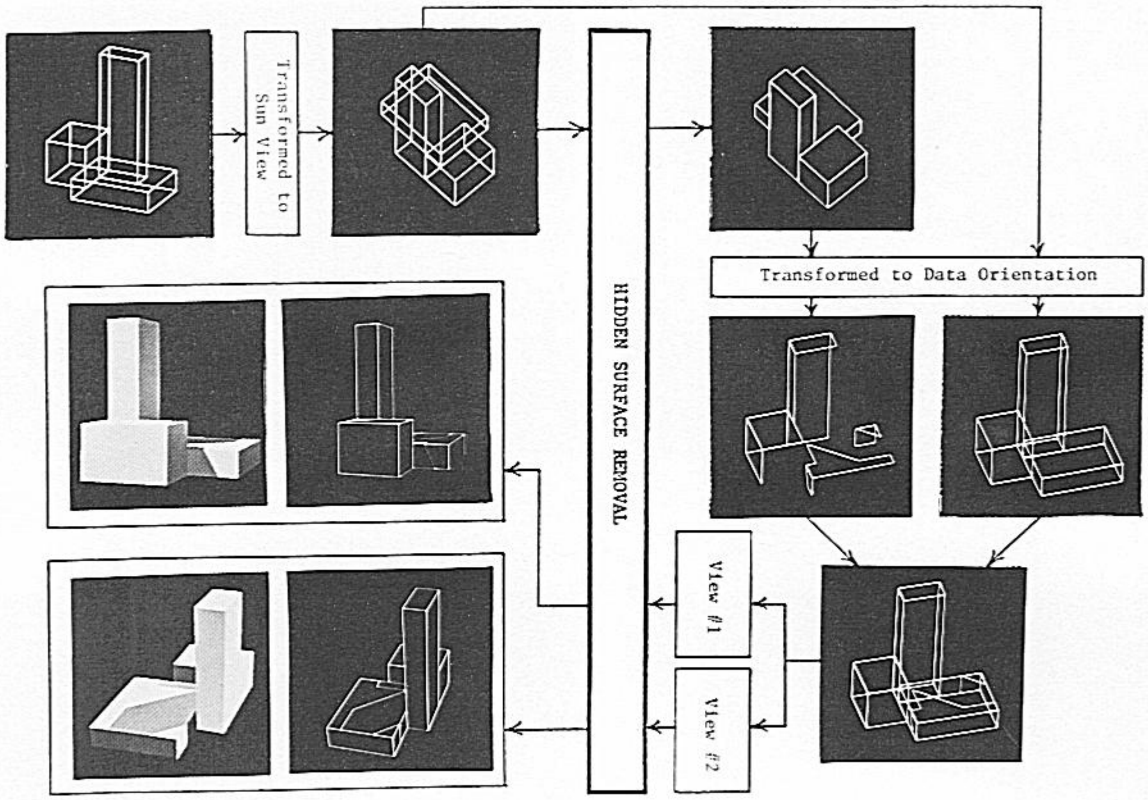


Figure 4. Shadow Creation and Display Process

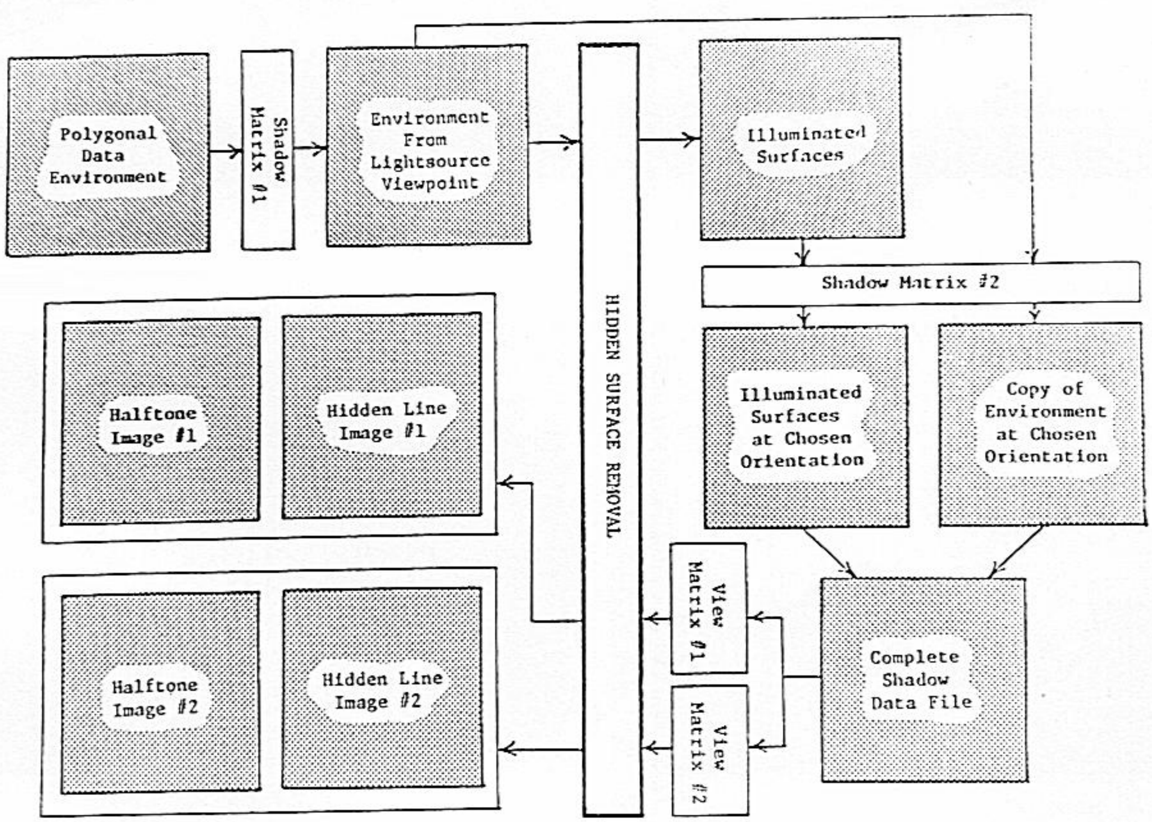
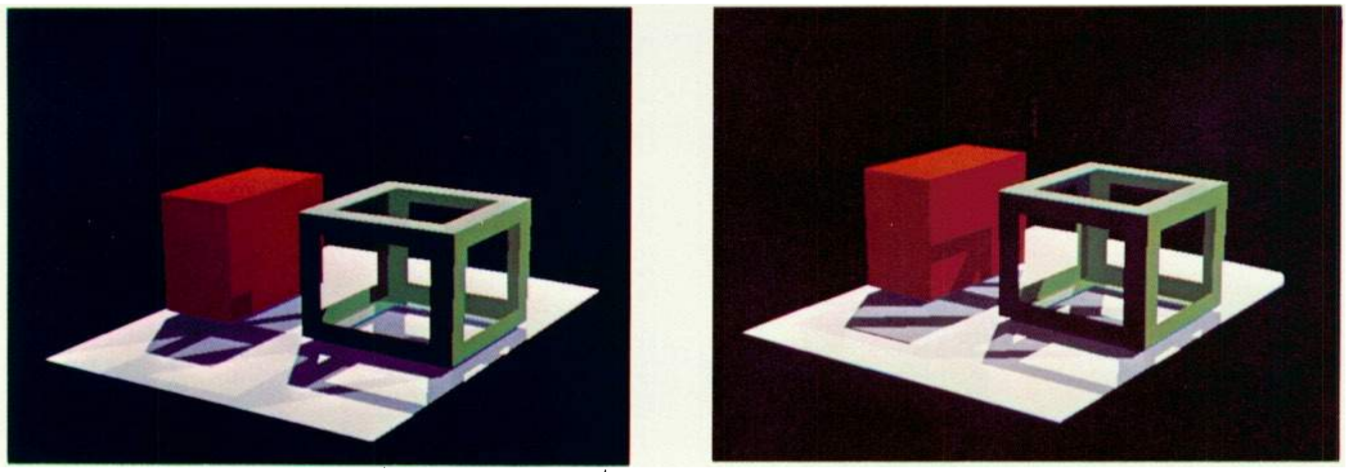
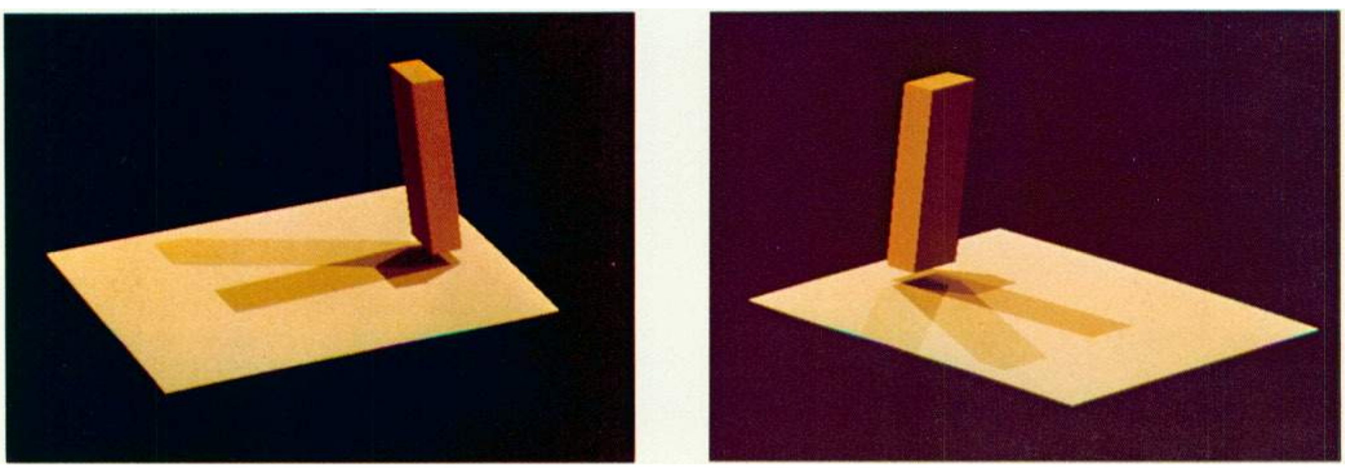


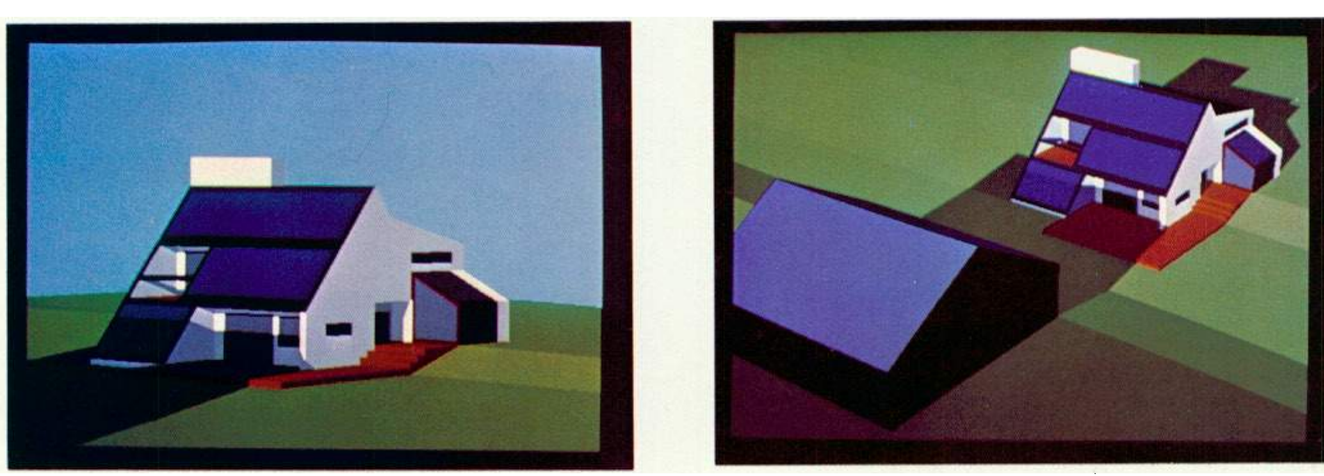
Figure 5. Shadow Creation and Display Process (key).



Figures 6. Shadowed Image Displays with Two Light Sources at Different Locations.



Figures 7. Shadowed Image Displays with Three and Four Distinct Light Sources.



Figures 8. Visual Examination of Simulated Shadowed Site from Two Observer Positions.

and blue) defining the color. The ratio of the primary hues used remains the same and only the intensity varies.

VII. SHADOW DISPLAYS WITH MULTIPLE LIGHT SOURCES

Shadowed images with more than one light source can be made by a single pass through the hidden surface removal program from each light source viewpoint.

Shadow data files may be utilized in the same manner as any other polygon data file that can be used by the hidden surface removal system since the same coordinate polygonal data structure has been maintained. The shadow data file may be transformed by a single matrix to any desired viewing orientation. It may also be used as a data file to which an additional set of shadows from another light source may be added (Figure 6 and 7). By maintaining the polygonal data structure throughout the hidden surface removal system, new possibilities in the area of graphic shadow generation such as the casting of shadows through translucent surfaces and the study of multiple reflections of light can be accomplished.

VIII. SHADOW APPLICATIONS

Since the object space computations are carried out to the limits of machine precision, numerical analysis of shadow areas is quite accurate and can be used for such important energy problems as measuring the effect of shadows on available solar energy or the reduction of air conditioning load. For a full yearly analysis, these effects need to be evaluated on an hourly basis, requiring an immense amount of computation time for complex environments.[9] Methods are presently being developed to predict hourly shadow areas based on the calculated results for a limited number of known solar positions.

Used as a graphical tool, designers may visually "walk through" a simulated site and see exactly where the shadows may fall at various times and on different days (Figure 8). This visual technique is much more effective than numerical output. The actual locations of shadows become obvious, and possible solutions to architectural design or lighting problems are readily apparent. A designer may interactively move or change a structure and continuously examine the results.

IX. CONCLUSION

The polygon shadow generation approach presented is both accurate and versatile. It would seem apparent that with the continuing advances in display hardware, the production of higher resolution raster scan programs will become more time consuming since their computations are directly related to the display resolution. Thus, it is important to refine this polygon approach in terms of speed and portability as its acceptability and use should certainly increase.

Shadows clearly enhance the depth perception and realism of computer generated images but several challenges remain. One problem still existing is the specification of the path through color space which should be used to depict the color changes due to shade and shadows. This aspect is particularly important when dealing with multiple light sources. Another problem involves finding a method for the modeling of light sources of finite size, such that the cast shadows contain an umbra and penumbra. Each of these problems are subjects for future investigations.

ACKNOWLEDGEMENTS

The research is part of a project sponsored by the National Science Foundation under grant number DCR74-14694 entitled "Development of Computer Graphics Techniques and Applications." The authors wish to particularly thank Ted Crane and David Bessel for their work in the implementation of the color display system.

REFERENCES

1. Appel, Arthur, The Notion of Quantitative Invisibility and the Machine Rendering of Solids, Proceedings ACM, 1967 National Conference.
2. Atherton, Peter, Polygon Shadow Generation with an Application to Solar Rights, M.S. Thesis, Cornell University, February 1978.
3. Bessel, David H., Crane, Ted, Programmer's Guide for Release I, Evans and Sutherland Video Frame Buffer Programming Package, Manual, Program for Computer Graphics, Cornell University, February 1977.
4. Bouknight, W. Jack, A Procedure for the Generation of Three-Dimensional Half-Toned Computer Graphics Presentations, CACM, Vol. 13, No. 6, September 1970.
5. Bouknight, W.J., Kelley, Karl C., An Algorithm for Producing Half-Tone Computer Graphics Presentations with Shadows and Movable Light Sources, 1970, SJCC, AFIPS, Vol. 36.
6. Crow, Franklin C., Shadow Algorithms for Computer Graphics, Proceedings of 4th Annual Conference on Computer Graphics, Interactive Techniques and Image Processing - SIGGRAPH 1977.
7. Kelley, K.C., A Computer Graphics Program for the Generation of Half-Toned Images with Shadows, M.S. Thesis, Univ. of Illinois, 1970.
8. Nishita, Tomoyuti and Nakamae, Eihaitro, An Algorithm for Half-Toned Representation of Three-Dimensional Objects, Information Processing Society of Japan, Vol. 14, 1974.
9. Rogers, Richard, Atherton, P., Nall, Daniel and Greenberg, Donald P., A Means for Including Shadowing in a Building's Thermal Analysis, CAD '78.
10. Weiler, Kevin, Hidden Surface Removal Using Polygon Area Sorting, M.S. Thesis, Program of Computer Graphics, Cornell Univ., Jan. 1978.
11. Weiler, K. and Atherton, P., Hidden Surface Removal Using Polygon Area Sorting, Proceedings of 4th Annual Conference on Computer Graphics, Interactive Techniques and Image Processing, SIGGRAPH 1977.