6. REFERENCES

1. T. Okoshi, **Three-Dimensional Imaging Techniques**, Academic Press, New York, 1976.

2. M. Lucente, "*Interactive Computation of Holograms Using a Look-up Table*," **Journal of Electronic Imaging**, Vol. 2, No. 1, pp. 28-34, January 1993.

3. M. Lucente, P. St. Hilaire, S. A. Benton, D. L. Arias, and J. A. Watlington, "*New Approaches to Holographic Video*," SPIE Proceeding #1732, July 1992.

4. M. Deering, "*High Resolution Virtual Reality*," **Computer Graphics** (SIGGRAPH'92 Proceedings), Vol. 26, No. 2, pp. 195-201, July 1992.

5. C. Cruz-Neira, D. J. Sandin, and T. A. DeFanti, *"Surround-Screen Projection-Based Virtual Reality: The Design and Implementation of the CAVE,"* **Computer Graphics** (SIGGRAPH'93 Proceedings), pp. 135-142, July 1993.

6. K. W. Arthur, K. S. Booth, and C. Ware, "*Evaluating 3D Task Performance for Fish Tank Virtual Worlds,*" **ACM Transactions on Information Systems**, Vol. 11, No. 3, pp. 239-265, July 1993.

7. R.C. Gonzalez and P. Wintz, **Digital Image Processing**, Second Edition, Addison-Wesley, Reading, Massachusetts, 1987.

8. G. Wolberg, **Digital Image Warping**, IEEE Computer Society Press, Los Alamitos, California, 1990.

9. E. Catmull and A. R. Smith, "*3-D Transformations of Images in Scanline Order*," **Computer Graphics** (SIGGRAPH'80 Proceedings), Vol. 14, No. 3, pp. 279-285, July 1980.

10. A. R. Smith "*Planar 2-Pass Texture Mapping and Warping*," **Computer Graphics** (SIGGRAPH'87 Proceedings), Vol. 21, no. 4, pp. 263-272, July 1987.

11. G. Wolberg and T.E. Boult. "*Separable Image Warping with Spatial Lookup Tables*," **Computer Graphics** (SIGGRAPH'89 Proceedings), vol. 23, no. 3, pp. 369-378, July 1989.

12. S. E. Chen and L. Williams, *"View Interpolation for Image Synthesis,"* **Computer Graphics** (SIGGRAPH'93 Proceedings), pp. 279-288, July 1993.

13. S. J. Adelson and L. F. Hodges, "*Stereoscopic ray-tracing*," **The Visual Computer**, pp. 127-144, Springer-Verlag, 1993.

14. S. Laveau and O. Faugeras, *"3-D Scene Representation as a Collection of Images and Fundamental Matrices,"* INRIA, Technical Report No. 2205, February, 1994.

15. S. T. Barnard "A Stochastic Approach to Stereo Vision," SRI International, Technical Note 373, April 4, 1986

16. R.C.Bolles, H. H. Baker, and D. H. Marimont, *"Epipolar-Plane Image Analysis: An Approach to Determining Structure from Motion,"* **International Journal of Computer Vision**, Vol. 1, 1987.

17. O. Faugeras, **Three-dimensional Computer Vision: A Geometric Viewpoint**, The MIT Press, Cambridge, Massachusetts, 1993.

18. B. Horn and B.G. Schunck, *"Determining Optical Flow,"* **Artificial Intelligence**, Vol. 17, 1981.

19. B. Lucas and T. Kanade, *"An Iterative Image Registration Technique with an Application to Stereo Vision,"* **Proceedings of the Seventh International Joint Conference on Artificial Intelligence**, Vancouver, 1981.

20. D. F. Rogers, **Procedural Elements for Computer Graphics**, McGraw-Hill, New York, NY, 1985.

21. D. Anderson, *"Hidden Line Elimination in Projected Grid Surfaces,"* **ACM Transactions on Graphics**, October 1982.

screen's width by the horizontal screen resolution measured in pixels, and using this result to scale a unit vector in the direction of the screen's horizontal axis relative to the tracker's orientation. A similar calculation determines $\vec{v}$.

The system takes as input any number of left and right eye images and their corresponding disparity or range images. The program can operate in a variety of modes. Given a single reference image and a disparity image, the image warper can generate convincing stereoscopic views over a large operating range. The precise range is dependent on scene content, but, we are often able to move within regions that are ten or more times larger than the interpupillary distance. If multiple images are used, the image warper will first select the reference view closest to the current head position, and warp it into appropriate left and right eye views. This method can significantly increase the working range.

Tracker latency in combination with rendering time are the two most significant problems in our prototype system. These delays must be kept extremely low in order to sustain the illusion that an object's position is fixed in space. Our system is capable of generating monoscopic 512×512 images in excess of 10 frames per second. However, when used stereoscopically with images as large as 1248×482, the performance drops below 3 frames per second. This, combined with at least 30 milliseconds of tracker latency, causes a noticeable delay between the time when the head moves until the time when the appropriate images are displayed. These delays cause the apparent three-dimensional image to initially rotate away from the observer just prior to snapping into position. We call this phenomenon swimming, and we are exploring solutions.

4. CONCLUSIONS

We have presented a new approach for the transmission of three-dimensional display data that is capable of synthesizing images from a wide range of viewing positions. The algorithms presented use image data along with image flow to reconstruct alternate views. When coupled with head-tracking, this technique could be used to synthesize truly three-dimensional images which appear to be fixed in space. In a broadcast environment, the same image and flow data could be used to support multiple simultaneous viewers. This would require head-mounted displays and separate image-warping capabilities for each participant, but, the content of the broadcast data stream would be independent of the recipients.

We believe that images can be warped in real-time with lower-cost hardware than traditional computer graphics approaches. We can also imagine hybrid scene generation systems where traditional graphics hardware would be used to render wide field-of-view images and range data (a common by-product of most rendering methods). These images and range data could then be broadcast to multiple simultaneous users, each wearing headsets with dedicated image-warping capabilities. Such an approach might provide a solution for the conflicting goals of high-quality rendering and interactivity.

There are still many open issues concerning image-warping based scene generation. Among these are, improved techniques for the reconstruction of the irregular grid surfaces that result from the warping, the potential of combining warps from multiple reference images to fill in obscured image regions, and real-time systems for capturing optical flow information.

5. ACKNOWLEDGMENTS

We are indebted to the following individuals for the contributions and suggestions on this work: Henry Fuchs for his visionary insights and continuing support of this research, Andrei State for comments and evaluations of the stereo quality, Kevin Arthur for insights on fish-tank VR, Russell Taylor for development of the networked tracker libraries, Mark Mine for explanation of the device independent VR libraries, Donna McMillan for her help in proofreading and editing text, and all the members of the UNC/UPenn collaborative telepresence-research group.

This research is supported in part by Advanced Research Projects Agency contract no. DABT63-93-C-0048, NSF Cooperative Agreement no. ASC-8920219, Advanced Research Projects Agency order no. A410, and National Science Foundation grant no. MIP-9306208.

enumerations based on the projected coordinate (x', y') .

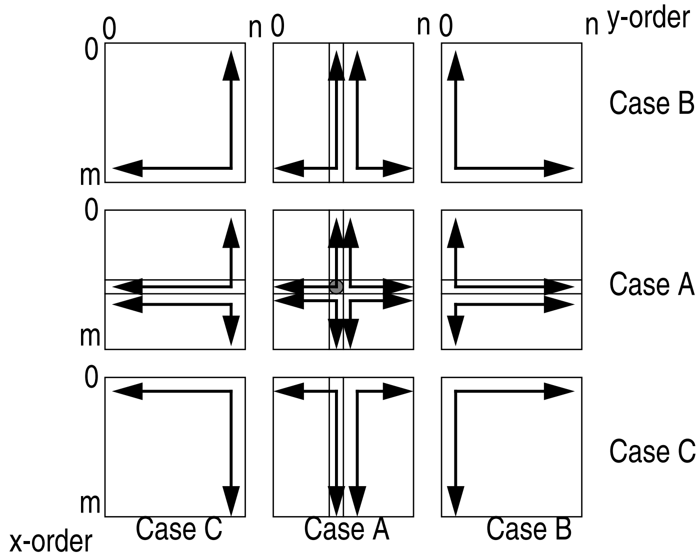

Figure 4. Facet enumeration order when the value of *w* is negative.

The enumeration order would typically be implemented as nested loops. Thus, the projection of the desired center of projection, (x', y') , would first be used to determine whether the surface is to be subdivided into one, two, or four sheets. Next, the starting value, terminal value and increment direction for each of the sheets is assigned. The facets are then painted by looping through the three-space coordinates of the grid points within a given sheet. These grid points are then reprojected to the desired view, V_x. The resulting image will be the correct hidden surface solution for the original projected grid surface when viewed from the new viewpoint.

3. RESULTS

We have built and demonstrated a prototype head-tracked stereoscopic display system that implements our image-warping based techniques. The software was written in C and operates in a standard UNIX environment as an X-windows application. The output of a high-resolution workstation display (1280×1024) is sent to a vertical frequency doubler which drives a Crystal Eyes, LCD-based, stereo glasses system on alternate frames. The stereo images are each computed at a resolution of 1248×482 pixels and displayed with the left image above the right image on the screen. Since the vertical-frequency doubling has the effect of stretching the aspect ratio vertically, the resulting images have square pixels.

The head's position is tracked by a Polhemus 3-Space electromagnetic tracking system whose transmitter is mounted overhead at a height of approximately 2 meters. The receiver unit is mounted just above the temple of the shutter glasses. The vertical offset of the receiver from the eye's position and the interpupillary distance between the eyes are entered as parameters to the program.

Our tracker interface library reports the position and orientation of the receiver, in the coordinates defined by the transmitter, in the form of a 4×4 matrix. From this matrix we extract position information and two orthogonal vectors, one along the line-of-sight and a second oriented along an axis parallel to the direction of the eye's relative displacement.

Since the projection screen is fixed in space relative to the transmitter, the projection parameters $\vec{o}$, $\vec{u}$,and $\vec{v}$ are determined by measuring the distance from the tracker to the screen's origin, as well as the screen’s height and width. The $\vec{o}$ vector for the left and right views is determined by subtracting the screen’s origin from the corresponding eye’s position. The $\vec{u}$ and $\vec{v}$ vectors will be the same for any head position. They are determined by dividing the

the i and j directions are considered independently. When the w element of e' is positive, the facets can be enumerated as follows:

- If $x' = x/w$ falls within the range 0 to $n-1$, then the list of facets must be partitioned into two groups; the first enumerated from $F_{0,j}$ to $F_{x',j}$ and the second ordered from $F_{n-1,j}$ to $F_{x'-1,j}$.
- If $x' = x/w$ is less than zero, then a single ordering from $F_{n-1,j}$ to $F_{0,j}$ is required.
- If $x' = x/w$ is greater than or equal to n, then a single ordering from $F_{0,j}$ to $F_{n-1,j}$ is needed.

The three cases are shown in Figure 3 with the eye's projected position shown as a darkened circle.

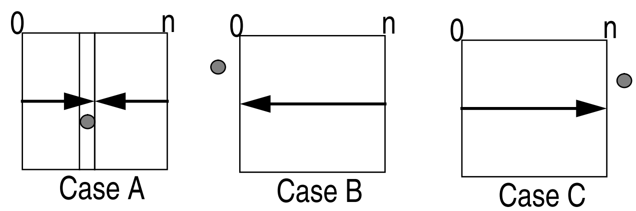

Figure 3. An occlusion-compatible enumeration order for x can be determined by projecting the desired center of projection onto the reference image's view plane. This projection is shown as a darkened circle.

Next, the facets are ordered in y according to the following three cases.

- If $y' = y/w$ falls within the range 0 to $m-1$, then partition the facets into two groups; the first ordered from $F_{i,0}$ to $F_{i,y'}$ and the second ordered from $F_{i,m-1}$ to $F_{i,y'-1}$.
- If $y' = y/w$ is less than zero, then a single ordering from $F_{i,m-1}$ to $F_{i,0}$ is used.
- If $y' = y/w$ is greater than or equal to m, the order is from $F_{i,0}$ to $F_{i,m-1}$.

Figure 4 shows the nine possible enumeration orders after both the x and y orders have been determined. All of the projected grid enumerations are shown relative to the single eye position shown in the center of the figure.

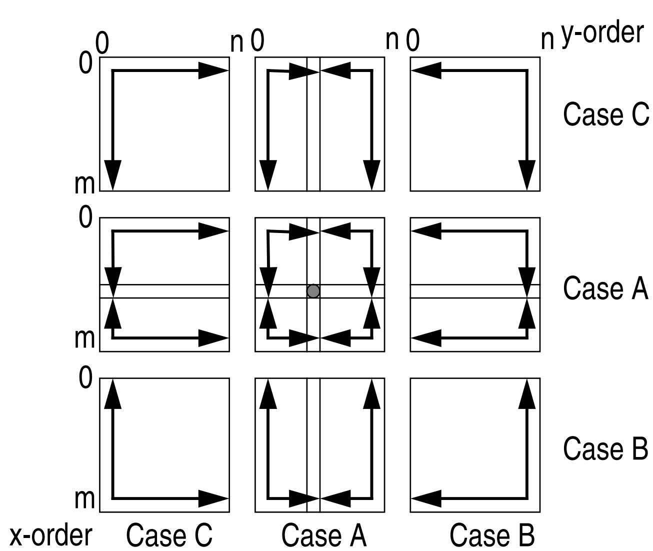

Figure 4. Facet enumeration order when the value of w is positive.

A similar breakdown results when the sign of w is negative. The following diagram shows the nine possible

standard image warping methods without any appeal to the implicit geometric content of the scene.

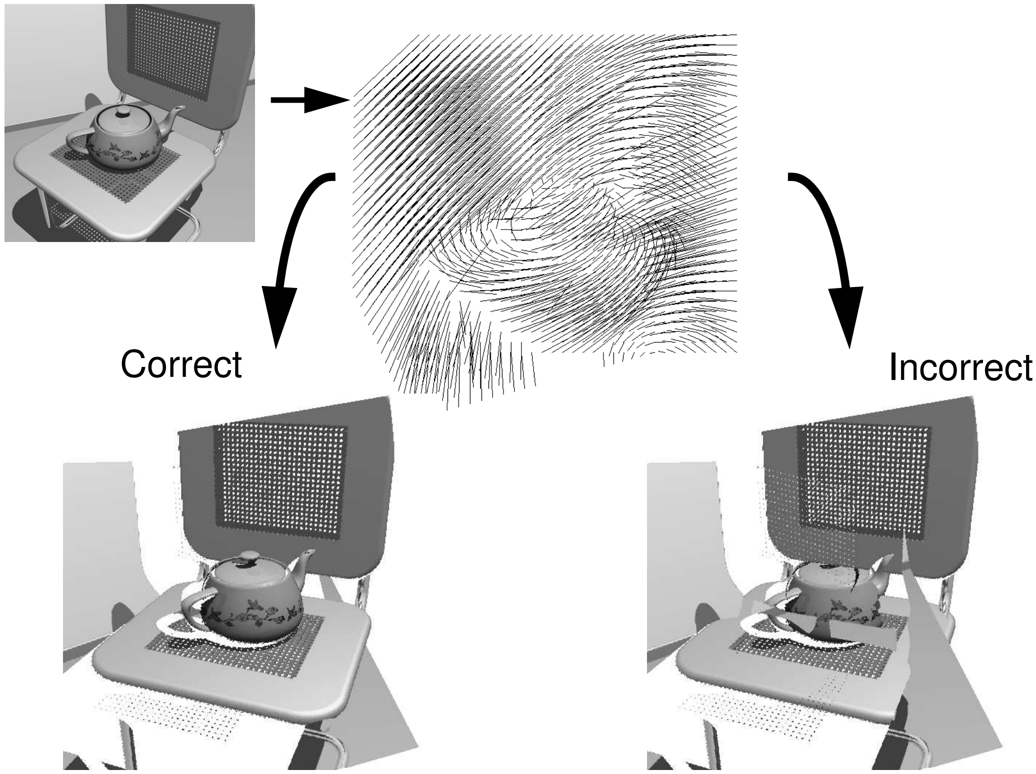

Figure 2. The image warp's mapping function allows invalid visibility solutions.

Our visibility algorithm considers reference images as connected topological meshs. The four neighboring pixels, (x_i, y_i), (x_{i+1}, y_i), (x_i, y_{i+1}), and (x_{i+1}, y_{i+1}) form a facet of the mesh. The desired algorithm will enumerate the facets $F_1, F_2, F_3, \ldots, F_N$, where $N = (n-1)(m-1)$, in such a way that if $\boldsymbol{F}_j$ occludes $\boldsymbol{F}_i$ then $j > i$. We will call this order *back-to-front occlusion compatible*. This approach is similar to Anderson's[21] algorithm for bivariate functions, with the exception that his visibility algorithm enumerates the facets in a front-to-back occlusion order. Anderson's choice of a front-to-back order requires that some representation of the grid perimeter be maintained to aid in deciding what parts or edges of subsequent facets need to be rendered. This representation requires auxiliary storage for maintaining a list of edges defining the perimeter. This list must then be queried before each new facet's edge is displayed, and the display must be updated if any part of the facet is visible. In contrast, a back-to-front ordering requires no additional storage because the proper occlusion is handled by the drawing order in a way similar to the classic painter's algorithm[20].

The algorithm for redisplaying a reference image from an alternate viewpoint, $\dot{c}'$, is described as follows:

1) Compute the projection of the desired center-of-projection onto the reference viewplane. This can be computed as follows:

$$x = (\dot{c}' - \dot{c}) \cdot (\vec{v} \times \vec{o}) \qquad y = (\dot{c}' - \dot{c}) \cdot (\vec{o} \times \vec{u}) \qquad w = (\dot{c}' - \dot{c}) \cdot (\vec{u} \times \vec{v})$$

2) Generate a proper enumeration order based on the values of (x, y, w).
3) Warp facets from the original projected grid to the new viewing position.

In order to generate a proper enumeration order based on the value of (x, y, w) several cases must be considered. In the following case analysis, $e' = (x, y, w)$ represents the projected viewpoint or eye position. The enumeration along

$$x' = \frac{ax + by + c\delta + k}{gx + hy + i\delta + m} \qquad y' = \frac{dx + ey + f\delta + l}{gx + hy + i\delta + m}$$

$$\vec{n} = \vec{u}' \times \vec{v}' \qquad \vec{r} = \vec{v}' \times \vec{o}' \qquad \vec{s} = \vec{o}' \times \vec{u}'$$

$$a = \vec{u} \cdot \vec{r} \qquad b = \vec{v} \cdot \vec{r} \qquad c = (\dot{c} - \dot{c}') \cdot \vec{r} \qquad k = \vec{o} \cdot \vec{r}$$

$$d = \vec{u} \cdot \vec{s} \qquad e = \vec{v} \cdot \vec{s} \qquad f = (\dot{c} - \dot{c}') \cdot \vec{s} \qquad l = \vec{o} \cdot \vec{s}$$

$$g = \vec{u} \cdot \vec{n} \qquad h = \vec{v} \cdot \vec{n} \qquad i = (\dot{c} - \dot{c}') \cdot \vec{n} \qquad m = \vec{o} \cdot \vec{n}$$

and the generalized disparity value is given by

$$\delta(x, y) = \frac{\|\vec{o} + x\vec{u} + y\vec{v}\|}{r}$$

Note the similarity between the generalized disparity expression and the classic expression:

$$\text{disparity} = \frac{\text{baseline} \times \text{focal length}}{\text{depth}}$$

This equation assumes parallel viewplanes for both reference images and that depth is defined along the optical axis. When this situation cannot be achieved mechanically, image rectification is used to satisfy this assumption. We can make direct comparisons between the standard disparity measure and the generalized notion. The baseline computation parallels the vector connecting the centers of projection used in the computation of the constants c, f, and i. The role of the focal length is equivalent to the length from the center of projection to a point on the viewplane, as seen in the numerator. Depth is replaced by the radial distance, or range value, from the center of projection. Traditional disparity values can easily be converted to the generalized format by dividing out the length of the baseline and multiplying by the tangent of the angle formed by the ray and the optical axis. The primary advantage of this generalized disparity notion is that it allows for arbitrary reprojections.

2.2. Visibility determination

While the image-warping function correctly determines the image coordinate of each pixel in the resulting projection, the possibility exists that it may also introduce many-to-one mappings, called topological folds. A situation is depicted in Figure 2 where two different images result from the same reference image and image-warp functions; the only difference was the order of evaluation. Ideally, only the front-most surface would be displayed. Determining the visible surface at each pixel position is one of the fundamental problems of traditional computer graphics.

One approach to solving this problem would be to treat the image as a spatial height-field and to use traditional computer graphics techniques to transform and scan convert the result. A standard Z-buffer algorithm could be used to determine the visibility on a pixel-by-pixel basis. There are problems with this approach, though. The transformation process requires the computation of an additional rational expression to determine the z-value for each pixel, and a screen-sized memory array is required to store the z-values. This approach is nearly the same as the transmission of a geometric database described earlier.

We have developed an alternative approach to determining the visible surface which does not require an explicit conversion to a geometric representation. This approach has the following important properties. It determines a unique evaluation order for computing the image-warp function such that surfaces are drawn in a back-to-front order; thus, it allows a simple painter’s style visibility calculation. It maintains the spatial coherence of the image, thus allowing the warp function to be computed incrementally. And, most significantly, the enumeration order can be computed independent of the disparity values for each pixel. This last property is surprising since all of the information concerning the geometry or shape of the underlying data is represented by this disparity information. Since the algorithm is capable of determining the visible surface using only the centers-of-projection of the reference and resulting images, along with the projection parameters of the reference image, this approach allows us to use only

Given the image and the pixel motion information, we will need a method for warping and determining visibility in the resulting image. The next two sections describe methods that are simple and efficient.

2.1. Image warping

The key components of any image warping technique are reference images and mapping functions. For autostereo applications we will require estimates of the relative spatial positions for the centers of projection of each reference image and an accurate camera model. The required mapping function must somehow capture the geometric nature of the underlying scene. Two possible candidates for this mapping function are the optical flow fields and the stereoscopic disparity fields. These two mappings are nearly equivalent in the case of a static scene, with their only distinction being the size of the baseline. This distinction is nontrivial. It has a dramatic impact on the underlying assumptions that can be made and on the resulting algorithm used to approximate a solution. We merely suggest that the underlying structure of the two vector fields is quite similar.

The image warping function that we use to generate images from new view points has the following general form:

$$x' = f(x, y, \delta)$$

$$y' = g(x, y, \delta)$$

$$I(x, y) \rightarrow I'(x', y')$$

where x and y are pixel coordinates of the reference image, δ is a generalized disparity value, *f ()* and *g()* are the mapping functions unique to the desired viewing position, and x' and y' are the pixel coordinates of the warped image. The mapping of pixel intensity values is indicated by the third equation.

In order to derive the mapping functions, we must first give a description of the projection model assumed. The planar projection model used in our image warping approach can be described by a point and three vectors, $\dot{c}, \vec{o}, \vec{u}$ and $\vec{v}$, as shown in Figure 1, where the point, $\dot{c}$, describes the center of projection, $\vec{o}$ denotes a vector from the center of projection to the coordinate of the origin of the viewing plane, and $\vec{u}, \vec{v}$ form a basis set for spanning the view plane. This formulation is capable of describing any planar projection. It naturally allows for the skewed and off-axis projections required for the display of stereo images on a single display screen.

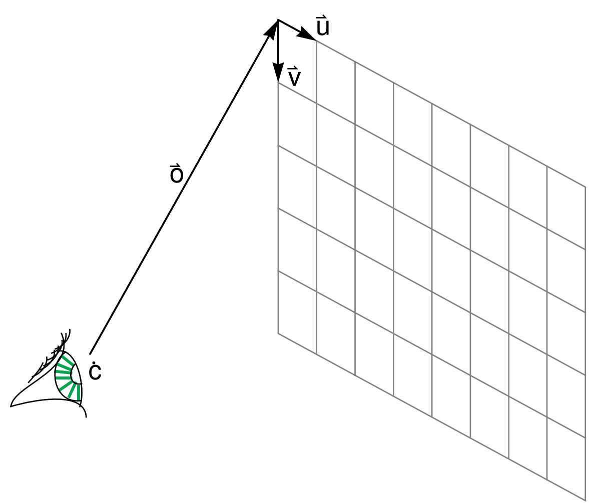

Figure 1. A point, the center-of-projection, and three vectors determine the planar projection.

Only the relative magnitudes of the vectors $\vec{o}, \vec{u}$, and $\vec{v}$ are significant. Thus, it is often convenient to scale them such that one or both of $\vec{u}$ and $\vec{v}$ are unit-length.

We can now specify the mapping functions *f ()* and *g()* as the following rational expressions:

rearranged points. Finally, this function is resampled to generate a final image.

Catmull and Smith[9] developed a technique for simplifying image-based geometric transformations into separable functions of x and y. This allowed the transformation, reconstruction, and resampling phases to be computed efficiently with scanline techniques. Subsequently, Smith[10] and Wolberg[11] have extended separable geometric transforms to include a wider range of transformations. These extensions are capable of describing any modification of an image by conceptually printing it on a rubber sheet, and then stretching and rotating it within the image plane.

1.3. Previous work

Other researchers have also explored the use of image warping for generating a wide range of views. Chen and Williams[12] described a technique called view interpolation. View interpolation employs reference images and pixel-mapping information to reconstruct arbitrary viewpoints with some constraints on gaze angle. Chen and Williams noted that one of the key difficulties encountered when using image warping to generate arbitrary new views was the determination of the visible surfaces. They chose to presort the quadtree compressed flow-field in a back-to-front order according to its z-value. This approach works well when all of the partial sample images share a common gaze direction, and the synthesized viewpoints are restricted to stay within 90 degrees of this gaze angle. Chen and Williams used pre-rendered synthetic images to determine flow fields from the z-values. In general, accurate flow field information between two samples can only be established for points that are mutually visible to both samples.

The methods used by Adelson and Hodges[13] to reconstruct stereo pairs from a single ray-traced image for static images and animations have some similarities to an image-warping approach. They refer to their approach as reprojection reflecting their treatment of the reference image as three-dimensional spatial sample points rather than a simple two-dimensional image. Adelson and Hodges also discuss an algorithm for enumerating the sample points which guarantees a correct visibility solution independent of the scene's content. This order is entirely dependent on whether a right eye is being reprojected to the left or vice versa. This visible surface approach is, however, applicable only to generating stereo-pairs. It would not work correctly for arbitrary viewing points and directions, as required to satisfy the motion parallax depth cue.

Another significant image-warping-like system has been described by Laveau and Faugeras[14]. Their approach recognizes that the epipolar geometries between images restrict the image flow field in such a way that it can be parameterized by a single disparity value and a fundamental matrix which represents the epipolar relationship. They also provide a two-dimensional ray-tracing-like solution to the visibility problem which does not appeal to an underlying geometric description. Their method does, however, require establishing correspondences for each image point along the ray's path. They also discuss the combination of information from several views, but primarily in terms of resolving visibility. By relating the reference views and the desired views by the homogenous transformations between their projections, Laveau and Faugeras can compute exact perspective depth solutions.

2. IMAGE-WARPING BASED SCENE GENERATION

The application of image warping techniques to broadcast television would allow viewers to see stereo images that are correct for their individual point of view. The broadcast signal would combine the conventional image with a second signal of comparable bandwidth that describes the how the pixels move as the viewpoint changes. The broadcast receiver would use the pixel motion information to warp the image to provide both the disparity required for stereo depth perception and motion parallax required to make the image appear stationary and solid when the viewer's head moves. Lower cost receivers might only provide for stereo display that is correct from a single view position, while conventional television sets would display only the image data.

The pixel motion data that is broadcast with the image signal is trivial to determine for the computer-generated synthetic imagery that is becoming increasingly common in program material, but it can be very difficult to determine from a set of reference images of a real scene. In our work we will make no particular claims about the suitability of the various methods for extracting image flow from scenes. In our experience, each method seems to have its own strengths and weaknesses. From this point on we will assume the existence of appropriate pixel motion data and will refer readers to the extensive literature on computer vision[15,16,17,18,19].

adjusting the focal length of our eyes over a narrow range in order to adjust this focus. The degree of this adjustment is monitored by the brain, and it is used as a third depth cue called accommodation. Finally, as we change the position of our head, significant information about three-dimensional structure can be determined by the changes that occur in the retinal image. This depth cue is known as motion parallax. Any display system must present to the user at least one of these four depth cues in order to create a sensation of three-dimensions.

Many three-dimensional display systems use stereopsis as the only depth cue by presenting separate images to each eye. In general, these images are correct for only a single viewing position and eye separation. At other viewpoints and separations, we interpret the stereo images as a slightly skewed and rotated three-dimensional object. Any motion of the observer that introduces a change in eye position will induce a perceived motion of the object being displayed. This apparent motion results from our attempts to mediate between conflicting motion parallax and stereopsis depth cues.

In order to produce a convincing illusion of a three-dimensional scene, we must either constrain the viewing position and, thus, prohibit head motion or, the system must be capable of displaying appropriate images that correspond to the changes in viewpoint.

1.2. Encoding three-dimensional information

A truly three-dimensional display system must be capable of generating images from a wide range of potential viewing positions. There are several possible methods for encoding the information required to generate these various views. One approach is to encode the images into a hologram as the summation of their constructive and destructive wavefronts. This approach can provide stunning results. It is unique in its ability to display three-dimensional data in a form which satisfies all four physiological depth cues, without any sort of viewing aid. Unfortunately, the transmission of holographic information requires tremendous bandwidth. This bandwidth requirement can be reduced somewhat by constraining parallax affects to a single dimension[3]. Other techniques have also been investigated[2]. In his classic survey of three-dimensional imaging technologies, Okoshi[1] (who dedicated most of the last chapter to speculations on the various forms of three-dimensional television) estimates that somewhere between 3 GHz (for horizontal parallax only) to 1500 GHz (for general-purpose wide-angle viewing) of communication bandwidth would be required to transmit holographic images at a frame-rate and quality level comparable to today’s television.

A second method of encoding view information is to transmit a geometric description of the scene. Standard three-dimensional graphics techniques could be used to transform, project, determine the visibility of, and render the geometric elements of the scene description for each eye. Standard stereoscopic display techniques could be used in conjunction with head tracking to view the apparent three-dimensional scene. Similar techniques are commonly used within the virtual-reality community[4,5,6]. The primary disadvantage of this approach is that it is generally limited to displaying synthetic imagery. With the introduction of high-speed, three-dimensional scanning technologies which simultaneously capture geometric and photometric information, there is some hope for overcoming this problem. A more severe difficulty, though, is that while today’s high-end scene generators approach rendering rates in excess of 5 million primitives per second, the geometric complexity of the real world may be as much as two-orders of magnitude greater. One trick that is often used to increase the apparent geometric complexity of a computer graphics scene is to apply texture on otherwise simple geometries. The skillful combination of textures and geometry can provide very convincing image generation capabilities.

A new and promising method of encoding a wide-range of view images for stereoscopic display, and a central topic of this paper, is the application of image-warping techniques for generating alternative views. Image warping is a rapidly growing branch of image processing which deals with the geometric transformation of two-dimensional images[7,8]. Geometric transformations modify the spatial relationships between pixels in an image and, therefore, are capable of image manipulations impossible to achieve by filtering alone. These manipulations include resizing, rotation, and skewing. Geometric transformations are characterized by mapping functions expressed as $x' = f(x,y)$ and $y' = g(x,y)$, where x and y represent the coordinates of a point within the source image, and x' and y' correspond to the coordinates of the same point in the result. This process can be logically subdivided into three tasks. First, a spatial transformation is applied which defines the “rearrangement” of all points within the input image on a new image plane. Second, a resulting intensity function is reconstructed by fitting a continuous representation through the

Head-tracked stereoscopic display using image warping

Leonard McMillan and Gary Bishop

University of North Carolina, Department of Computer Science
Sitterson Hall, Chapel Hill, NC 27599

ABSTRACT

In traditional stereoscopic displays, the virtual three-dimensional object does not appear to be fixed in space as the viewer's head moves. This apparent motion results from the fact that a correct stereo image can only be formed for a particular viewpoint and interpupillary distance. At other viewpoints, our brain interprets the stereo image as a slightly skewed and rotated version of the original. When moving the head, this skewing of the image is perceived as apparent motion of the object.

This apparent motion anomaly can be overcome with head tracking. Unfortunately, a head-tracked stereo-display system requires the generation of images from arbitrary viewpoints. This has previously limited the practical use of head-tracked stereo to synthetic imagery. We describe a stereoscopic display system which requires the broadcast of only a stereo pair and sparse correspondence information, yet allows for the generation of the arbitrary views required for head-tracked stereo.

Our proposed method begins with a pair of hyper-stereo reference images. From these, a sparse set of corresponding points is extracted. Next, we use image warping and compositing techniques to synthesize new views based on the user's current head position. We show that under a reasonable set of constraints, this method can be used to generate stereo images from a wide range of viewpoints. This technique has several advantages over previous methods. It does not require an explicit geometric description, and thus, avoids the difficult depth-from-stereo problem. We also describe a unique visibility solution which allows the synthesized images to maintain their proper depth relationships without appealing to an underlying geometric description.

Keywords: Computer Graphics, Picture/Image Generation, Stereoscopic Display, Image Processing, Image Warping, Image Morphing, Three-Dimensional Graphics

1. INTRODUCTION

It is intriguing to consider one day in the future when the technology may exist for the broadcast of three-dimensional television. Our high-expectations are already being set by the portrayal of such display systems in the various works of science fiction. The technological realities, however, present many obstacles to the realization of these fantasies. We have developed a unique scene generation method which we believe has tremendous potential for overcoming some of these obstacles. We start our explanation with a discussion of the phenomenon of three-dimensional perception, followed by a description of possible methods for encoding three dimensional information. Next, we discuss our method of scene generation which is based on image-warping techniques. We describe how this method can be used to generate stereoscopic images from a wide range of viewing positions. Finally, we present our results from constructing a prototype system based on our methods.

1.1. Three-dimensional perception

We perceive the world as three-dimensional primarily because of four physiological properties[1]. First, we combine information from two spatially distinct sensors (our eyes) into a single representation. The discrepancies between the separate images sensed by the eyes are resolved by the brain, which places a three-dimensional interpretation on them. This depth cue is called stereopsis. Second, when observing an object, we tend to rotate our eyes so that the optical axes converge at the point of observation; this depth cue is called binocular parallax. Third, since our eyes have a finite aperture size, we cannot see at differing distances in clear focus at once. Fortunately, we are capable of