ADAPTING TRADITIONAL MEDIA FOR VIRTUAL REALITY ENVIRONMENTS

by

David M. Geshwind
President
LATENT IMAGE DEVELOPMENT CORPORATION
Two Lincoln Square
New York, NY 10023
212/873-5487

INTRODUCTION

There is a great deal of 'hype' surrounding the field of Virtual Reality. It has been heralded as a paradigm shifter that will revolutionize everything from eduction to entertainment, medicine to sex. At that same time, we have heard that Virtual Reality is also (second only to nuclear energy, perhaps) the technology most likely open to misuse of dystopic proportions.

The hype aside, there is much to be excited about, and I am as enthused as anyone. As a baby boomer and a card carrying member of the first TV generation, I have grown accustomed to the ten-mega-bit-per-second rush of information that is television -- it takes a lot to get my attention. The experience of immersion and interactivity afforded by Virtual Reality is new enough and dense enough to be truly interesting, independent of program content. Eventually, however, this newness will wear off; content and usability will become the important aspects of VR applications.

With all the excitement, something basic is overlooked. The first 'alternate' reality that we have any record of are cave paintings in France and Spain, about 40,000 years old. The concept, that reality could be represented -- or even influenced -- by image, probably had a more profound impact on the culture at that time than VR will have today.

From this start, imaging technology has advanced periodically, but the basic concept has remained the same.

IMAGING TECHNOLOGY	APPROXIMATE AGE
Cave Paintings	40,000 years
Perspective	600
Photography	160
Motion Pictures	100
Film Sound	65
Full Color Cinema	55
Stereoscopy (3-D)	Comes and goes
Dolby Stereo	20
Digital SPFX	15
Interactive Media	15

Now Virtual Reality has added the quality of immersion.

But, rather than an entirely new technology, VR is an incremental advance. Just the latest in a progression of visual communication techniques.

As such, VR can benefit from a time tested practice, that of adaption from other media. Recent film history is rife with adaptions from earlier works.

In the 1940s Lawrence Olivier directed, stared in and received the Oscar for several highly praised adaptions of Shakespeare's plays. Three years ago Kenneth Branagh was Oscar nominated for doing the same thing with Shakespeare's "Henry V"; only this time with location shots, stereo sound and special effects. Shakespeare's plays have also provided the basis for films as varied as the science fiction classic "Forbidden Planet" (from "The Tempest") and the musical "West Side Story" (from "Romeo and Juliet"). Shakespeare himself often adapted from history.

In other examples: Civil War documentaries 'zoom and pan' to create a cinematic effect from old black and white photographs; Kenny Rogers has adapted no fewer than four made-for-TV movies from a single song, "The Gambler"; and, in general, films, television programs, and even video games are adapted from books, comics, songs and personalities, as well as from each other.

"Star Trek" is a self-referential labyrinth -- the television program that begat the Saturday morning cartoon, scores of books and magazines, six theatrical films and, eventually, another television series "Star Trek: The Next Generation". An interesting example comes from that new series. The 'holodeck' is the ultimate VR system, providing full visual immersion, tactile feedback and freedom of motion. It is often prominently featured in the program, which is set in the 24th century. But the only uses of the holodeck shown for literary purposes involved Shakespeare, again, Sherlock Holmes and "Dixon Hill" a doubly fictitious Sam Spade clone circa 1930. Even within the extreme creative freedom afforded by writing fiction about the far future, the inhabitants of that future (or at least the television writers) do not invent new works for their media, but adapt the old.

Taking a lead from the past (and the fictional far future) practitioners of VR might do well to consider if, how and when existing media may be adapted for the mew medium.

There are several advantages associated with adaption from another medium, as opposed to starting from scratch. First, the content has been time tested and works -- though it may be changed or built upon during adaption. Second, presenting familiar elements often gives the viewer a higher level of comfort and a sense of continuity. For example, the MacIntosh 'garbage can' (for dispose or discard) and other icons have made using PCs much less foreboding for the computer-reluctant. Third, familiar elements also provide a way to communicate a wealth of implicit detail and context, based upon a prior experience shared by author and viewer. Fourth, at the early stages of a new technology, adaption provides a way to demonstrate what the new medium does well, by showing how a known property is changed, hopefully for the better.

The adaption of works to the VR environment will be different than earlier adaption processes. For example, with earlier adaptions, it was usually the story (or the equivalent organizational structure of non-fiction works) that was preserved, while the particular visual embodiment was discarded, to be re-made. With VR, we suggest the opposite will hold true.

Particularly because of the interactive nature of the VR medium, the story or other structure may have to be modified or extended, to take into account alternative possibilities derived from viewer choices. The element that may be more useful to adapt literally, is the imagery. Therefore, by modifying existing visual material, for incorporation into the VR environment, adaption can provide another advantage -- texture and detail.

Virtual worlds (the VR environments) tend to be very schematic; lacking the texture and detail that denotes 'real' reality. A recent feature covering VR on the ABC news program 20/20 was illustrated with some stunning, highly complex and subtle computer animation. But the feature was highly misleading because this animation was presented as typical of VR. On the contrary, these few minutes of animation represented months of design and programming, and many days of intensive computation. There is no way that the current real-time simulators that are the core of VR systems could produce such images.

Although VR's bare bones look is partly due to (current, but surmountable) computational limitations, it is also extremely hard to design and specify virtual worlds in great detail.

With computer programming and CAD-like modeling functions being the two ways of creating the data bases that are the computer's versions of these worlds, specifying reality is an arduous task. By adapting existing media -- primarily film and video, but also photographs, illustrations and paintings -- we believe that a new level of realism can be brought to VR, and a great deal of effort saved.

Certainly, to be useful, the adaption process must be less work than creating a VR environment from scratch. But, the use of the adaption process must also be appropriate. For example, one of the most frequent VR applications is to simulate architectural spaces or industrial machines, to determine human usability, prior to building or manufacture. When simulating a specific hypothetical space or object, adaption of an existing, but different, such item would be useless. But, many educational, training, scientific, medical and entertainment programs exist that are a rich source of material for adaption to VR applications.

At Latent Image, we have a patented process, StereoSysthesis,^{tm} to convert 2-D material to 3-D. We believe that our technology provides the basis for the integration of existing film, video, photographs, paintings and illustrations into Virtual Reality environments.

THE 2-D TO 3-D CONVERSION PROCESS

There is a long history of converting media from one format to another; films are routinely transferred to videotape, paintings photographed, photographs printed, and images are digitized into computer data bases.

However, in some cases, additional information, that is not explicit in the original format, needs to be synthesized. For example, two separate sound channels are synthesized from old monophonic recordings to create modern 'stereo' recordings. More recently, it has become common to add color to black and white films and television programs. In fact, we have a patented and improved process for film colorization, and our 2-D to 3-D conversion process is an outgrowth of that basic technology.

The two systems share the same 'front end', a human/computer collaboration on the problem of identifying which pixels (picture elements, or image dots) belong to which objects in an image frame. Even with the current state-of-the-art in artificial intelligence and pattern recognition, computers today are not up to the task of identifying arbitrary objects in arbitrary settings that change every few seconds; i.e., people and props in the scenes of a film. As it turns out, humans are still the best 'programmable devices' available for this pattern recognition task. But, computers are much faster and more accurate at the separation and repetition process. So we have redesigned the boundary and interface between the human and computer tasks in the process.

There are basically three approaches to the problem. In the first, every object in every frame must be painted, or at least outlined. This actually dates back almost 100 years where, in France, it was common practice to hand tint each object in every frame of every print (there was no way to print color films even from a color master) of short novelty films. This task was performed by row after row of women working at large magnifying glasses, using tiny brushes and transparent inks. The current state of film colorization is almost the same; although the hand painting is now carried out on computer and video systems, and once a master tape is created it can be copied.

An alternate, but only theoretical, approach is to have a human operator paint a first frame and have the computer paint all the other frames of a scene from this guide. Unfortunately, as explained above, computer technology is not now up to that task. The company that holds a patent to that approach, in fact, hired dozens of operators, who worked at rows of workstations, painting every object in every frame.

Our approach is a compromise which lets the human and the computer specialize in that part of the task at which each excels. We have a human operator outline the objects in two separated frames, usually between three and ten frames apart, depending upon the complexity of the motion in the scene. By bounding the process at both ends, there is no requirement for the computer to recognize, analyze, or even to reference the intervening frames. The computer is programmed to draw a progression of outline images, changing the shapes and positions of the outlines for each object in each frame. Starting with the first frame, the outlines are updated progressively in a series of steps spaced out over the entire set of frames, and eventually reach the final frame in the sequence. This process is called interpolation, in mathematics, or inbetweening, in animation.

Inbetweening is shown in Figure 1: where a tall diamond on the left of the screen would be outlined in Frame 1; a wide diamond on the right in Frame 9. Frame 5, the middle frame in the sequence, would then automatically be drawn with a diamond that is neither tall nor wide, and which is mid-way in the frame left to right. The other frames in the sequence would similarly be automatically drawn as a geometric progression.

When colorizing, the color overlay of the background, which is hand painted, is much more detailed than the overlays for the foreground objects, which are computer interpolated. This is contrary to the practice of the other facilities colorizing films. However, we have determined that, since the background is still (compared to the moving foreground objects) and the viewer's eye/brain has a long time to absorb the image, detail in the background is even more important than in the foreground to the perceived reality of the scene. (We suggest that practitioners of VR consider this effect, when designing VR environments.) We can afford to spend the time needed to add background detail, because the same background painting is used for each frame in the scene, and we can amortize the extra human effort over many frames.

Just as each object in a scene can be marked with a color tag for automatic coloring, each object can be marked with a depth tag for the creation of 3-D (stereoscopic) images. Our process can create 3-D for any display system and format available, including the dual screens as are used in VR helmets. Initially, however, we have been working with field sequential video and LCD-shuttered glasses.

Each of the thirty frames-per-second of video is composed of two fields. The even lines are displayed first, followed, 1/60 second later, by the odd lines. By placing the left-eye image in the even lines and the right-eye image in the odd lines (or, vice versa), we alternate display of the left and right images. LCD glasses are flickered, with one lens opaque and the other clear, and are reversed every 1/60 second. In this way a continuous stream of left images arrives at the left eye, and right images at the right eye.

Now, as to how the left- and right-eye images are made from an initial 2-D image. The principle of parallax dictates that as an object is moved further from or closer to the viewer, the object appears to be offset in opposite directions (left or right) in each eye. This is due to the lateral separation of the two eyes in the head. The parallax effect is synthesized by our system. After the 2-D image is separated into image parts, two composite images are built up from those parts, with each part relatively offset by an amount corresponding to its distance from the viewer, and in a direction corresponding to which eye the view is for.

The process is diagrammed in Figure 2. An original 3-D scene, consisting of a diagonal ruled background, a white circle in the mid-ground, and a black square in the foreground, is depicted at the top. Directly below is a 2-D image of that scene as if photographed, which is then separated into a left and right image pair further below. Note that the mid-ground circle is relatively offset slightly in each of the two images, and the foreground square is relatively more offset. Below that the left and right images are combined for display, which is finally viewed through 3-D glasses and the original 3-D scene reconstructed in the viewer's eye/brain complex.

However, in practice, the process is more complicated then described above. When parallax shifting image parts, to create the left and right views, holes can develop in the image. This is shown by a diamond (Figure 3A) which is shifted to create one view, either left or right, leaving a hole, shown by horizontal lines (Figure 3B). The hole may be filled by image material not of the original image. The 'patch' may be derived from an equivalent part of another frame, either earlier or later in the scene. Alternately, if a patch cannot be found from another frame, material can by synthesized by a human operator using a computer paint system, or by the computer extrapolating from what material is visible. (If material cannot be found in another frame, then there cannot possibly be a conflict between the missing material, as would have been shown in another frame, and the synthesized patch.) This patch is represented by vertical lines (Figure 3C). In an entirely different approach (Figure 3D), if the hole is created by moving and object closer to the viewer, the object may also, reasonably, be made larger, and hopefully cover the hole.

Another variation on the basic process, which is covered by our patent, is 'auto-tracking'. Once an object (again depicted by a diamond, Figure 4A) has been identified, a depth is selected and left and right offsets computed (Figure 4B). Next, the size of the object in a subsequent frame (Figure 4C) can be measured by the computer and, using calculations involving perspective and the characteristics of the camera lens, a new depth can be derived. The computer can then automatically calculate and effect a new set of parallax offsets (Figure 4D).

Until now, it has been assumed that objects are separated and shifted as if 'cardboard cut outs' (as if viewed from above, Figure 5A). Our patent also covers more sophisticated techniques. For example, the plane of objects can be tilted with respect to the screen (again shown as if viewed from above, Figure 5B) by specifying a depth at each edge. An even more sophisticated effect can be achieved by 'sculpting' each separate object (Figure 5C). Off-the-shelf systems exist, such as the SONY System G, that permit operators to distort a planar object, with live video mapped onto it, into a sculpted surface. In demonstrations, that system has, in real time, taken a flat image of a woman's face and turned it into a convincing sculpted 'mask'. In our system, left and right pairs of similarly sculpted image elements will produce an extremely realistic effect exhibiting not only depth, but complex shape and shading.

The above describes our basic 2-D to 3-D conversion technology, which makes it possible to incorporate existing 2-D photographs, illustrations, film and video as 3-D elements into VR displays. However, if full immersion into, and interactivity with, a film scene is what is desired, the process may be extended in the following way: First, note that the creation of the left and right images constitutes a slight change of the point of view between the eyes. By adding equivalent (rather than opposite) offsets to the left and right images, the point of view of the observer may be moved further, permitting interactive travel within the scene space. Second, when changing the point of view drastically, or rotating around the scene space, in addition to holes developing in the image, the reverse sides of image parts may become visible; for example, the back of someone's head. The same technique used to 'fill holes' can be used to either borrow material from another frame or scene, or to completely synthesize the reverse sides of objects. (See Figure 5D, where the dotted portion represents the hidden and synthesized reverse side.) In this way, entire virtual worlds may be created for each individual scene of a film, within which interactive scripts can be played out.

3-D PAINT SYSTEM

As a by-product of software created to implement the 2-D to 3-D conversion process, we have developed a '3-D paint system'. It is derived from the functions provided to the operator to fill holes created when image parts are parallax shifted. The system is currently a partially completed prototype of what will eventually be a self-contained, stand-alone product. Currently the software runs on an ATVista display board from Truevision, Inc., installed in an IBM-PC/AT.

The system is designed with 101 planes, 50 into the display and 50 out into room space, as well as the plane of the screen. Currently, a plane is selected and painting functions occur on, or relative to, that fixed plane. Later, with the addition of a 3-D input device, it will also be possible to change the selected plane dynamically.

The user interface consists of reasonably familiar menus of buttons, color and brush palettes, etc. In addition, a 'ruler' device is included which permits the selection of a particular depth as well as keeping track of those planes already used and those in protected mode. However, a major difference between this interface and that of standard paint systems is that all the graphic controls are in 3-D: all buttons move in and out and cast drop shadows; the ruler and message windows are deep openings into the basic menu plane, which in turn floats above the screen; cursors move in and out in the 'Z' direction, as well as in the X and Y directions; and, window region markers have depth as well as width and height.

Typical painting and graphic functions are provided which work on a single plane at a time. Other functions, for example some edge filtering, span more than one plane; and, facilities are provided to pick or exclude any one plane, or ranges of planes.

Several display modes are provided: with the left image displayed above the right image, useful for testing purposes; field interleaved modes with both left/right and right/left options operating at 60Hz for operation with LCD-shuttered glasses from 3DTV Corporation; and, left/right and right/left modes for 120Hz operation with StereoGraphic Corporation's Crystal Eyes product.

The 3-D paint effect consists of two major elements. The first element -- parallax offset -- can best be seen with the display in testing mode with the left and right images separated on the screen. (See Figure 6, for a schematic representation of this effect.) As can be seen, when the depth is set into the screen and painting (of the word 'FAR') is done, the horizontal position of the cursor and painting function is offset to the left in the upper image, the left-eye view; and, it is offset to the right in the lower image, the right-eye view. When the depth is set to out of the screen (the word 'NEAR') the opposite situation exists with the left-eye view offset to the right, and the right-eye view offset to the left. When the depth is set to the plane of the screen (the word 'MIDDLE'), the offset is identical in both views. The bunching of objects in the right eye view is due to the placement of the far object at the left of the screen, and the near object at the right.

The second element of the 3-D effect -- proper obscuring of objects at different depths -- would be seen as painting progressed at screen depth, over the marks previously made both in and out of the screen. As would be seen, middle depth painting covers up the marks that are further away, but marks that are closer are not covered up. This is possible because as each pixel is changed on the screen, not only are red, green and blue values placed in the frame store, but a Z or depth value is also stored, although not seen. Then, when the same pixel is a candidate for changing again, the depth value stored at that pixel is compared with the depth being painted. Only if the current pixel's depth is further away than that being painted, is it covered up.

This effect is incorporated into all 3-D paint functions. For example, when a complex 3-D image is brought into the display where another complex 3-D picture already exists, various elements of the two images properly intersect in 3-D screen space, with nearer objects obscuring those to the rear.

Once a 3-D input device is integrated into the system, other functions will be provided to permit 'sculpting-like' functions and truly 3-D objects will be able to be built up and painted.

This system provides an important adjunct to previous 3-D imaging tools. Prior software systems provide two basic methods to create 3-D objects. First, objects may be described by typing text at a keyboard, essentially a computer programming task. Second, other software packages exist which provide CAD-like functions where objects may be built up from lines, polygons or other graphic primitives and then modified, copied, etc.; some are even called '3-D Paint' systems. However, as far as we know, our stylus or mouse operated, pixel-oriented, stereoscopic paint system provides the only true painter's or illustrator's interface to the 3-D creation process.

CONCLUSION

2-D to 3-D conversion provides the basis for adapting existing media to a form usable with Virtual Reality systems. Extensions to the basic process will permit the creation of full immersion environments from films and video programs. Existing content structures may need to be extended to accommodate interactivity. However, the main advantage of adapting prior works to VR is that, incorporation of existing images makes the creation of Virtual Worlds easier, and the resulting environment richer and more detailed.

The 3-D paint system provides a new interface, for painters and illustrators, which allows a new approach to creating objects and environments for Virtual Reality applications. It also permits a new class of artists to access the VR creation process. We believe that broadening the creative community involved in VR will enhance the artistry of the entire field. This will facilitate not only the creation of more realistic virtual environments, but also, will enable the realization of fantastic personal visions.

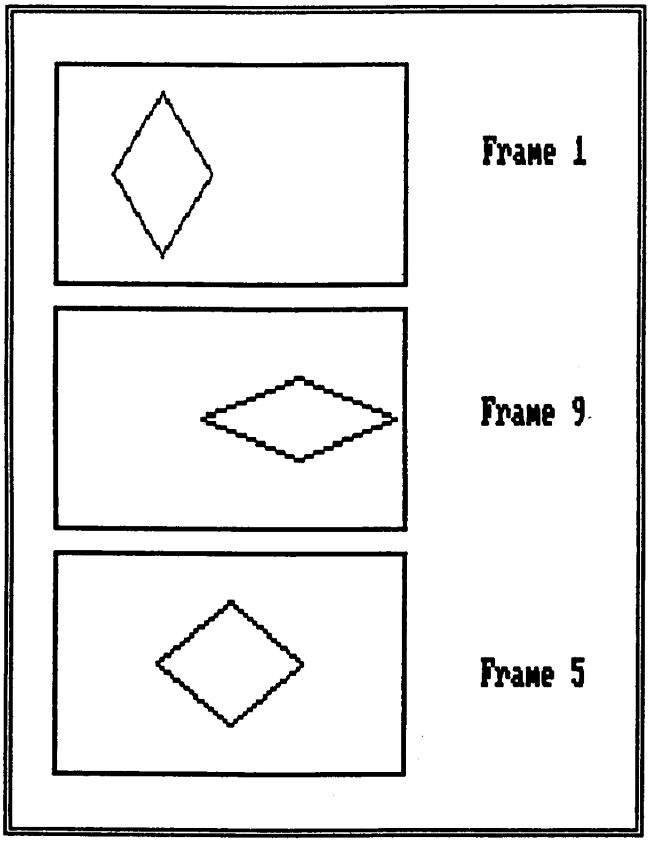

Figure 1: Interpolation or Inbetweening

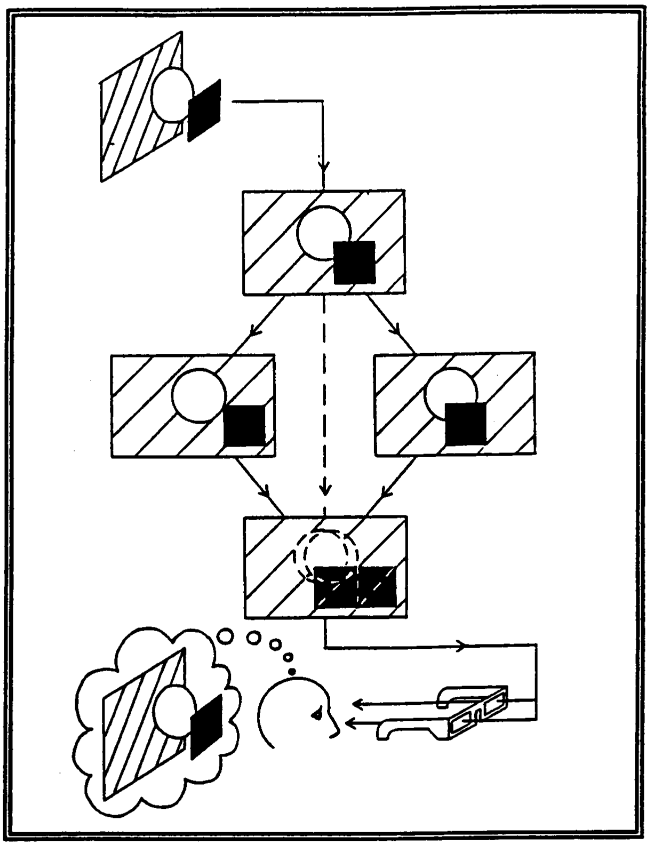

Figure 2: 2-D to 3-D Conversion Process

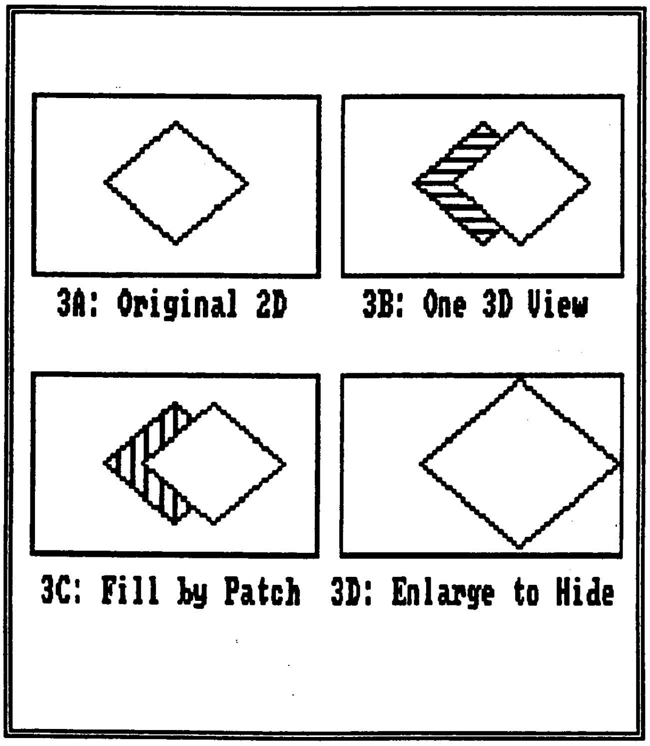

Figure 3: Creating and Filling Holes

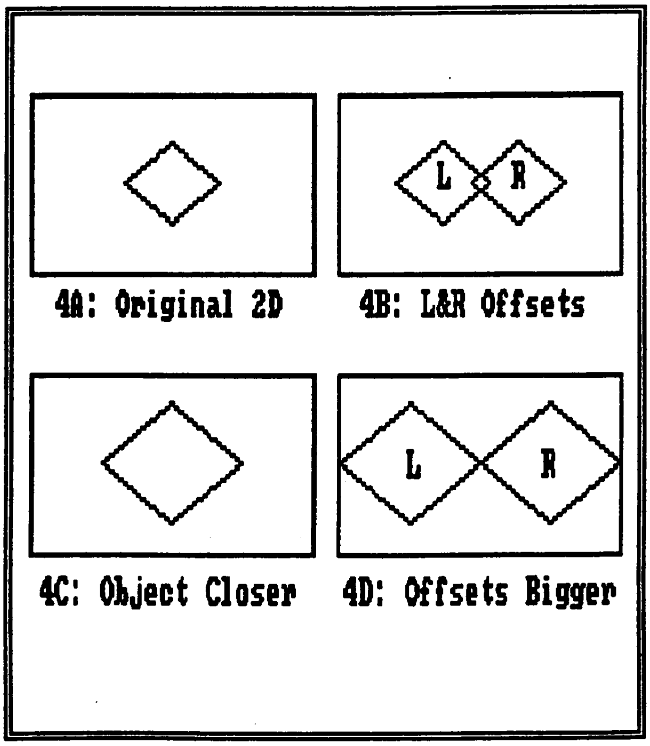

Figure 4: Automatic Depth Tracking

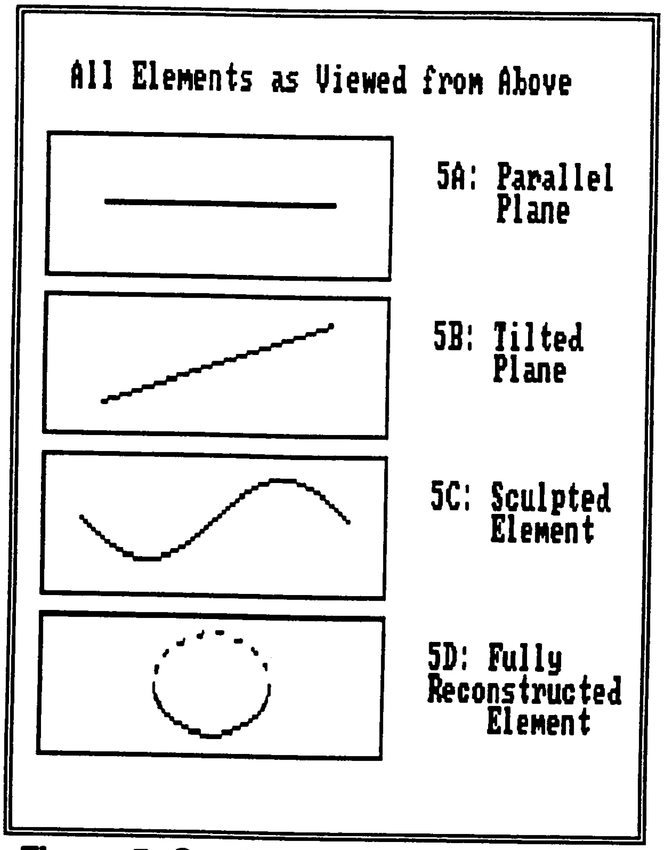

Figure 5: Synthesized Element Shapes

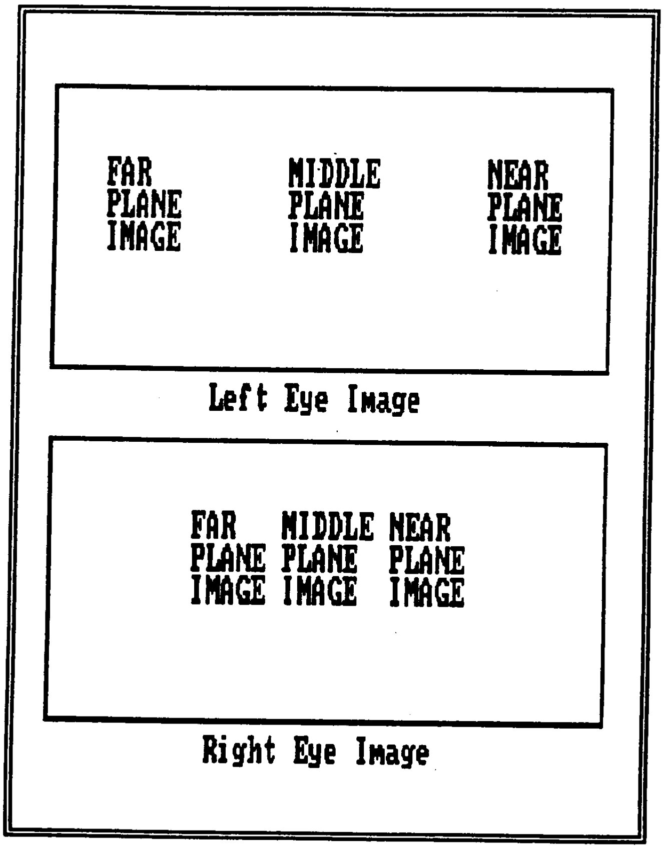

Figure 6: 3D (Stereoscopic) Paint System