ELEMENTS OF REALSPACE IMAGING

TECHNICAL
REPORT

APPLE
MULTIMEDIA
LAB

SAN FRANCISCO

Foreword

I met Michael Naimark in 1977, while visiting MIT to discuss a project that later became the Aspen Moviemap. He set up a crazy contraption on the floor, which I soon recognized as a scale model of a 360 degree wraparound movie screen. Inside it was a film projector on a rotating Lazy Susan. He turned on the projecter, and it spun slowly around, spewing out images on the screen. He looked at us and said, "Cameras move... Why not projectors?"

Since then, Michael and I have worked on a number of projects together, and he continues to push at the edges of our cultural and technological assumptions, particularly in the realm of cinematography and what he calls "realspace imaging". He prods me to wonder why images have to be constrained to rectangles, or flat surfaces. He challenges my thinking as I try to convey spatial information using computer and imagery technologies.

He has been key to many productions at the Apple Multimedia Lab, and an enthusiastic participant in both discussion and debate, particularly in our development of the Visual Almanac™. His knowledge of cameras and videodiscs, and his conceptual approaches to production work, have been invaluable.

Interestingly, I have directed the main work of the Multimedia Lab more toward the development of "conceptual spaces", information environments that represent content to interactive viewers. Yet as we move our work into networked worlds, where video and other data-rich elements are becoming readily available, Michael's development of alternative virtual environments becomes increasingly relevant. Clearly, these virtual worlds need not match the "real world"; but I believe a knowledge of their potential is extremely useful if we are to create information environments that are both comfortable and engaging.

So I asked Michael to write this paper. I hope it can assist you in challenging assumptions and conventions in the use of media, the way Michael has assisted me in our conversations over the last 15 years.

Kristina Hooper Woolsey

Kristina Hooper Woolsey
Director, Apple Multimedia Lab

Table of Contents

Acknowledgements

I fear that I've tormented my friends and colleagues hacking out this paper, and I am grateful to all of them. Everyone at the Apple Multimedia Lab was involved at some point, but particularly Fabrice Florin, Janey Fritsche, Steve Gano, Nancy Hechinger, Lisa Koss, Bob Mohl, Pat Roberts, Kristee Rosendahl, Rob Semper, and Nick West. Rachel Strickland, a colleague affiliated with the Lab, was also very helpful.

Phil Agre, Doug Crockford, Scott Fisher, and Brenda Laurel are all pushing the "realspace imaging" envelope in their own ways, and I am grateful for their time and energy.

I also wish to thank Catherine Boyle and Cindy Rink for their editorial input on this report, and Art Jagonasi for his fine illustrations.

Finally, a special note of gratitude goes to Kristina Woolsey, the Apple Multimedia Lab Director, for her continuing intellectual support.

Abstract

Along with the marriage of motion pictures and computers has come an increasing interest in making images appear to have a greater degree of realness or presence, which I call "realspace imaging." Topics such as high definition television, 3D, fisheye lenses, surrogate travel, and "cyberspace" reflect this interest. These topics are usually lumped together, with the implicit assumptions that "the more resolution, the more presence", and "the more presence, the better". This paper proposes a taxonomy of the elements of realspace imaging and, through examples of art and design, shows that quality is not proportional to bandwidth but is a complex interaction between form and content.

Preface

The evolution of this paper began with the appearance of the videodisc in the late 1970s, whose significance, as Andy Lippman once said, is that it "sanctified the frame." In those early days, we often spoke about the fact that videodiscs could play movies, stillframes, and "any stuff in between." MIT's Aspen Moviemap, produced in 1978-79, was such an exercise: we shot driving down every street in Aspen at a rate of one frame every ten feet, using a fifth wheel to trigger the cameras. Not coincidentally, Kristina Hooper Woolsey was involved with Aspen as well, then a psychology professor at U.C. Santa Cruz studying spatial representation and cognitive mapping.

Over the next ten years, Bob Mohl and I found ourselves travelling the world outshooting each other with this "stuff": Bob shot thousands of faces and registered them by eyes; I (along with my CalArts students) shot all the stars on Hollywood Boulevard in a similar way; Bob (along with John Borden for Bank Street College) moviemapped Polenque by shooting along walking trails; both Bob and I together (for the Paris Metro) moviemapped a section of Paris shooting along sidewalks; I (along with Ken Carson for the Exploratorium) moviemapped San Francisco shooting from a helicopter; Bob (along with Margo Nanny at the Apple Multimedia Lab) carefully shot sliced fruit in registration; I (along with my San Francisco Art Institute students) shot dishes of food for projection on a plate.

This "stuff in between" movies and still pictures could simply have been undersampled realtime movies, but it wasn't. The sequences we were shooting were naturally "browsable." They rarely had any significant time direction, and were just as browsable backward as forward, and at any speed. They were "duration-independent".

They were also perceptually continuous, since there were no "cuts" during sequences. When you moved forward or backward, you experienced a natural, seamless visual continuity.

Sound familiar? Just like "reality." These two features, duration independence and perceptual continuity, allow browsable movies to reproduce the way we navigate through the real world. They are qualities we do not experience through conventional linear cinema. Though browsable movies need not be explicitly spatial (like moviemaps), browsing virtual spaces via videodiscs is a natural application.

It couldn't happen at a better time. Over the last decade, the computer graphics world has developed and refined 3D visual databases, where a 3D virtual "space" can be stored and accessed by a free-moving "virtual camera." Advances in technologies allow us to move through such worlds in real time: the user wears special goggles, fitted with stereoscopic wide angle displays, over each eye. A position sensor allows the user's head movement to control the virtual cameras.

Videodiscs employ real-world imagery shot with cameras. Graphics computers use "artificial" imagery, since inputting camera-generated images has not been sufficiently perfected. But videodisc-based browsable movies and computer-based 3D graphic databases share something unique and exciting: a sensibility that the display screens are windows into another world, a world that may be as real as reality itself.

This is where we begin.

ELEMENTS OF REALSPACE IMAGING

Michael Naimark

Introduction

I have often wondered what it would take to make me believe there really is a little man inside the box when I'm watching the Evening News.

Well, for one thing, I'd have to believe little people existed.

I would also have to be convinced that what I was seeing was indistinguishable from reality, that it was a *window* rather than a video screen. It's well documented that the first audiences of the Lumiere Brothers' cinema ducked at a scene of an oncoming train.[1] We don't, which implies that at least part of our perception of "reality" is acquired.

Using media to create a sense of being somewhere else is not new. An early term for such a concept is "telepresence", coined by Marvin Minsky in the 1950s.[2] Various other terms have been used, including "microworlds",[3] "artificial reality",[4] "virtual environments",[5] and "cyberspace".[6]

I define the term "realspace imaging" as the process of recording and displaying sensory information indistinguishable from unmediated reality. Imagine looking at a framed image as if it were a window. Fooling the eye into believing the image is real is a difficult task. Fooling two eyes is even more difficult. Fooling two eyes while allowing the freedom of head motion is more difficult still.

Now imagine having the freedom to move around. Now add time-based phenomena such as motion and sound. These are the elements of realspace imaging I want to discuss.

One well-known example of such a "mediated" experience is the flight simulator, where a single user is surrounded by a believable image, the cockpit physically moves in sync with that image, and perhaps most importantly, the user is in control.

Three-dimensional and panoramic movies, video games, videodisc-based surrogate travel, and high definition television have spawned further interest in "being there" through media. Head-mounted displays, where the user wears stereoscopic, wide-angle, position dependent video displays, were first demonstrated in 1968[7], but have been enjoying recent popularity.[8]

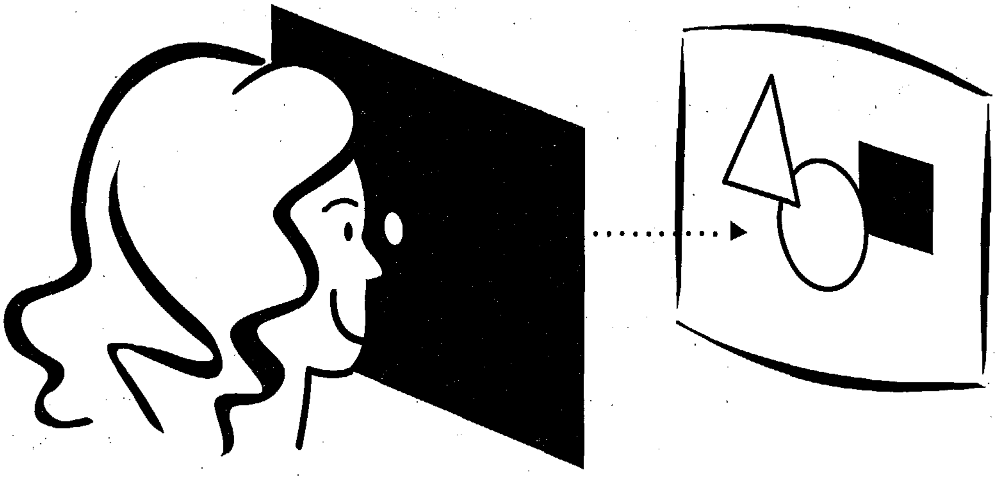
Three relatively new, unrelated industries are converging here: multimedia (computers plus video), special effects (computers plus film), and "virtual reality" (an entirely bizarre collection). In each case, the intent is to convey a strong sense of "place."

"Place" is a funny word. We use it both literally and metaphorically. We often blur the distinction between the two, particularly when dealing with anything virtual, which is, by definition, when dealing with media. Thus we talk about "navigating through information spaces" the same way as we talk about finding our way to the airport.

This report was written as much from a cinematographer's point of view as from a computer person's. Computer people working with generated 3D imagery have the luxury of having "virtual cameras" free to move at their command anywhere in their virtual world. Cinematographers need real cameras. Historically, many of the realspace imaging issues have been pioneered with real cameras. Also, as 3D modelling increases in sophistication, the need for real cameras as digitizers of the real world becomes more apparent.

I have added examples and exceptions that are my personal favorites. In a sense, the exceptions are most interesting. Although they seem to cancel the value of the element in question, this is not my intent. Rather, I wish to emphasize that maximizing all elements is rarely appropriate or necessary, let alone realistic.

I. Looking Through A Window From A Single Point Of View: Monoscopic Imaging

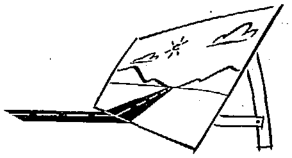


Imagine sitting in front of a wall with two frames on it. One is a real window showing the scene outside, the other contains a film or video image. That we can always distinguish between the two is significant. The same cannot be said of audio; most of us have mistaken a voice from a high fidelity radio as being one actually present in the room.

The eye is apparently harder to fool. This section addresses monoscopic imaging: viewing an image with one eye and from only one point-of-view, like looking through a single peephole.

Both eyes are even harder to fool. The next sections will discuss stereoscopic imaging, which requires two distinct points-of-view, and multiscopic imaging, where the eyes have local freedom of movement representing multiple points of view.

Monoscopic (or monocular) imaging represents a single point of view. Cameras always shoot from one point of view at a given instant. With very few exceptions (discussed later), virtually all images we see are monoscopic, whether they are in print, on a television, or in a movie theater.



Sometimes, such a constraint can be turned into a feature. For example, "forced perspective" is a Hollywood technique for continuing a foreground scene back by building a model or painting a backdrop which lines up perfectly, but only from the single point of view of a stationary camera. From any other point of view, the scene looks discontinuous.

Orthoscopy (Scale)

The most apparent difference between a framed image and a window is whether or not things are the right size. One is reminded of a classic Picasso story: When accused of distorting reality in his artwork, the artist asked the accuser if he had a girlfriend. He did, and produced a wallet-size photograph of her. Picasso replied, "She's beautiful but she's so tiny!"

Orthoscopic images, images viewed in proper scale, require the viewer to see from the same angle of view as that of the camera lens. For a given focal length, orthoscopic viewing requires a screen size dependent on the distance from the viewer. The greater the viewing distance, the larger the screen required.

To view a single image orthoscopically, one must view it through one eye with no head movement. Though orthoscopic imaging for single users is possible using a peephole or a head-mounted display, it is difficult for group viewing, unless all viewers are equidistant from the screen and near the axis of projection. Otherwise they would see the frame distorted from a rectangle into a trapezoid, as we do when we sit to the side or near the front of a movie theater.

Exceptions:

Practically every image we see is non-orthoscopic. It is remarkable how good we are at "reading" non-orthoscopic images. Nor do we have much problem viewing a movie from the side of the theater, even though we do actually see a trapezoid. Studies have shown that an extra degree of cognitive image processing is required to "straighten out" a trapezoidal image.[9] Ironically, we don't seem to mind much.

The zoom lens is another noteworthy exception. A zoom is where the orthoscopy changes: objects appear to get larger or smaller. Realworld "zooms" happen in our head, not in our eyes, since the actual retinal size of the image doesn't change. (An orthoscopically correct zoom would be where the projector lens zooms in sync with the camera lens.)

Spatial Resolution and Color

Spatial resolution may be the most-discussed aspect of realspace imaging. Today, the images we see on television are recorded on a wide variety of formats whose principal difference is spatial resolution (and color, partly a subset of spatial resolution and partly a subset of imaging and display technology). These include video 8 and VHS (both high band and regular), 3/4", beta, 1", D1 and D2, 16mm film, and 35mm film. Movie theaters often boast of 70mm film release prints because of more spatial resolution. The current dilemma over "how much resolution is enough?" has been given additional fuel by arguments over standardization for high definition television.

Normal NTSC video has a 525 line display (about 480 lines constitute the picture area). The NHK High Definition Television standard has an 1,125 line display. Research in this area suggests that a display with a 45 x 90 degree field of view would require a 3000 x 6000 pixel display for the lines to go away — many times the resolution of even the HDTV standard.[10]

Exceptions:

The marked preference we show for high resolution seems constrained to the area of motion pictures and photography. It is not an obsession we carry over into the arts.

Very few western painters, especially since the advent of photography, have based their work on the premise "the more resolution, the better". For example, it would be absurd to suggest that J. M. W. Turner wanted more detail in his large-scale paintings of storms at sea!

Eastern art, on the other hand, has always respected the concept of unresolved space. As early as the eighth century, Japanese Sumi-e painters based their landscapes on emptiness as much as image, allowing large blank areas on the silk or rice paper to convey distance between foreground and background. The most important element of Sumi-e painting was "knowing when one had 'said' enough".[11]

Spatial Consistency and Spatialization

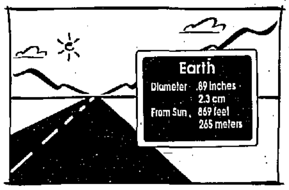
Many kinds of lower-resolution images appear acceptable, particularly if the noise or artifacts have realworld analogs, like video "snow" and real snowfall. But when the styles of resolution are inconsistent with each other, the image looks "wrong." Steve Yelick once referred to this as the "Gumby in New York" problem.

Video overlay usually exhibits this problem. An overlay can be "spatialized" by making it appear as if it is a contiguous part of the image, through lighting and shadow, scale and synchronous movement with the background.

Examples:

The use of bluescreen in film and chroma-key in video allows the opaque compositing of an independent foreground and background. The goal is usually to make the composite image look believable. We see this every day with our tv weatherperson standing "in front" of the weather map (which often is bubbling with dynamic effects), when he or she is actually standing in front of a blue-painted wall.

Films such as *Song of the South* and *Who Framed Roger Rabbit* successfully integrated graphic and cinemagraphic inconsistencies through the use of a simple narrative vehicle: by folding the inconsistency into the story and make-believing that graphic characters exist in the real world.



Another approach to compositing is spatialization of information; taking data or graphic material and giving it "realworldness" by placing it in space. Spatialization was used in *The Visual Almanac's* "Planetary Highway." Overlays were made consistent with the background imagery by turning them into road signs, literally giving them perspective in space (through video effects technology).

Exception:

An engaging example of spatial inconsistency is in a rock video of "Sting" and his band performing "Set Them Free." It appears as a straight documentation of the band performing the song, but each of the bandmembers has an intentionally different "look": Sting has a diffused halo around him, the saxophone player is in black and white, another moves in coarse movements, and the back-up singers artificially speed up and slow down. All the while the actual background looks "normal." The result is a subtle effect: "There is something wrong with this picture." Such subtlety doesn't come easy or cheap; each performer had to be carefully shot separately, the image processed, then mixed together.

Dynamic Range and Brightness

Dynamic range is the span from the whitest whites to the blackest blacks in visual recording and display. The eye has a broad dynamic range allowing us to simultaneously see bright outdoor scenes and shadow detail. Film has less dynamic range than the eye. Scenes shot in film must be carefully lit to "squeeze" the dynamic range into film's limits, such as "fill lighting" the shaded areas more and "key lighting" the bright areas less. Video has even less dynamic range than film, requiring more complicated lighting to achieve the same effect as film. Thus, film shot and transferred to video results in a greater dynamic range than video-originated material, which is half of the reason why many cinematographers prefer the "film look." (The other half involves frame rate, see below.)

Although dynamic range is more or less relative, brightness is the absolute upper limit of whitest whites, measurable by a light meter.

Example

I once took a light meter reading of an outdoor scene off the screen of New York's Ziegfield Theater, one of the highest quality theaters in the world. I then took a reading of a similar scene outdoors. The difference was an astounding ten f-stops, or a thousandfold difference in brightness. Remember, in a movie theater you are sitting in the dark, pupils wide open, even if you're looking at an image of a sunny day. The room is dark, of course, because the image is projected large over a wide area. A tradeoff exists between brightness, screen size, and directionality of the light off the screen.

Monoscopic Depth Cues

Monoscopic depth cues are the ones we can see with only one eye and no head movement, such as when we look through a peephole.

Content Cues

There are many cues to depth perception when the viewer's head is motionless that are based on image content, such as perspective, overlapping or occlusion, aerial perspective or atmospherics, light and shade, and textural gradient.[12] It is noteworthy that these cues are automatically captured with a camera. However, this is not the case for computer-generated imagery, where these factors have historically been of great concern.

Accommodation (Focus)

With one eye open and a fixed head position, a prominent depth cue is accommodation, the focusing of the eyes' lens by the surrounding muscles. It is similar to focusing the lens of a camera. There are two ways to determine depth from focusing a camera. One is to focus on an object and read the calibrated focus setting. We sense our eye muscles in a similar way: we can "feel" in our eyes when we focus on something close, for example. The other way to determine depth is by the amount of blur that exists for objects out-of-focus, which is partly a function of distance and partly a function of brightness. Obtaining depth data in an image by comparing two samples of blurriness has been demonstrated.[13]

Accommodation discrepancies are often most prominent while viewing landscape images, where the eyes should be focused on infinity (and where parallax diminishes to near zero). One method of achieving "infinity focus" is to project farfield imagery far away on a large screen, but this requires space. Another technique uses a large concave mirror that magnifies and refocuses an image from a small monitor. But even with this large mirror, the effective viewing area is still small, enough for only one person. Smaller and less expensive optics can be used in place of a mirror if the viewer is further restricted to a peephole.

A small concave mirror may be used for nearfield imaging by projecting a real image in front of the mirror's surface. The "floating nickel" illusion and video "genies" hovering in space are popular examples.

Example:

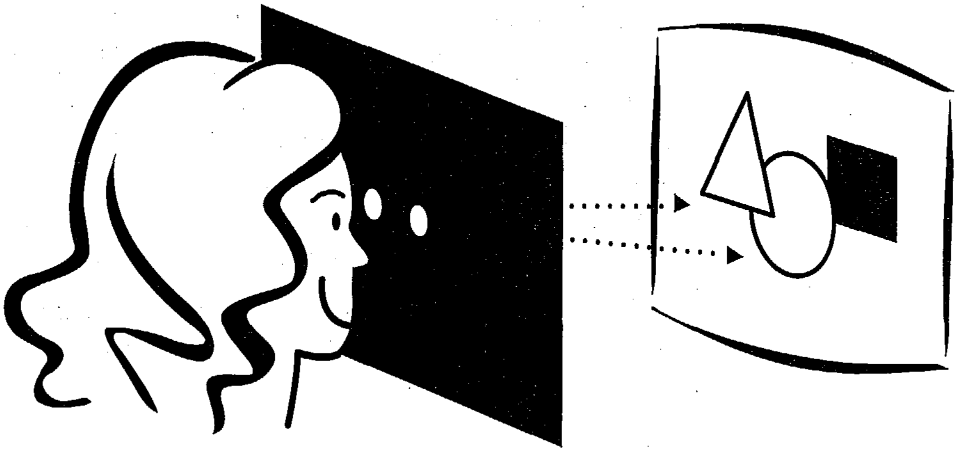
The flight simulation industry achieves infinity focus through a large concave mirror mentioned above. The goal is to allow both of the pilots' eyes to be focused at infinity, just as they would when they look out of a cockpit and everything is far away.

Exceptions:

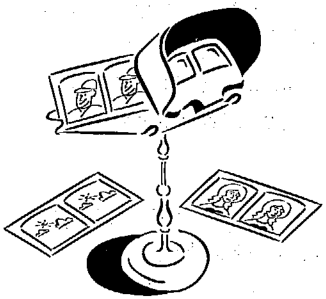
Like orthoscopy, most images we view are not properly accommodated, particularly if they contain material at varying depths. Consider the internal ambiguity of sitting six feet from a movie screen and viewing an image of a farfield landscape, where the closest object may be hundreds of feet away. The eye's accommodation muscles tell the brain "distance equals 6 feet" but the mind sees an image much further away.

A relationship exists between accommodation and brightness, since the brighter the image the greater the depth of field of the eye. Since displays have relatively low brightness, it may well be that brighter displays will minimize the need to explicitly address the issue of accommodation.

II. Looking Through A Window From Dual Points Of View: Stereoscopic Imaging



Stereoscopic (or binocular) imaging represents two single points of view, one for each eye, separated to give a noticable lateral displacement, or parallax. Parallax is often erroneously pitched as all that is necessary for depth in imagery (the stereoscopic movies of the 1950s were simply labelled "3D"). There's no easy way to record and reproduce parallax. First, two simultaneous views need to be recorded, with care taken for proper convergence and disparity. Then, each view must be seen exclusively by each eye.



Stereoscopic imaging has a lively history dating back at least to Wheatstone's invention of the stereoscope in 1833.[14] The most popular techniques require glasses to be worn (such as anaglyphic, polarized, or shutter). Methods not requiring glasses usually require that the head be held in a particular position (using mirrors, peepholes, or lenticular screens, for example).

At the very least, stereoscopic imaging requires two points of view, usually displaced by the nominal interocular distance of 65 mm. Greater distances produce an exaggerated or hyperstereoscopic effect, while reversing left and right produces a bizarre "inside-out" or pseudoscopic effect. For a virtual camera in a 3D database, accessing these two views is simply a matter of programming. But for realworld cameras, two views must be independently recorded. This may be done by a stereoscopic camera built for such tasks, which has at least two lenses and takes at least two pictures simultaneously. Or it may be done by moving a single camera laterally over time (which can be casually done by hand, often with passable though not perfect results). Ideally, everything must stay still since any kind of motion will counteract the stereoscopic effect.

Examples:

Perhaps our most popular experience with stereoscopy is with a "Viewmaster" viewer, where stereoscopic pairs (shot with a special Viewmaster camera) are individually displayed to each eye. It is small, cheap, and effective, but it is also made only for individual viewing.

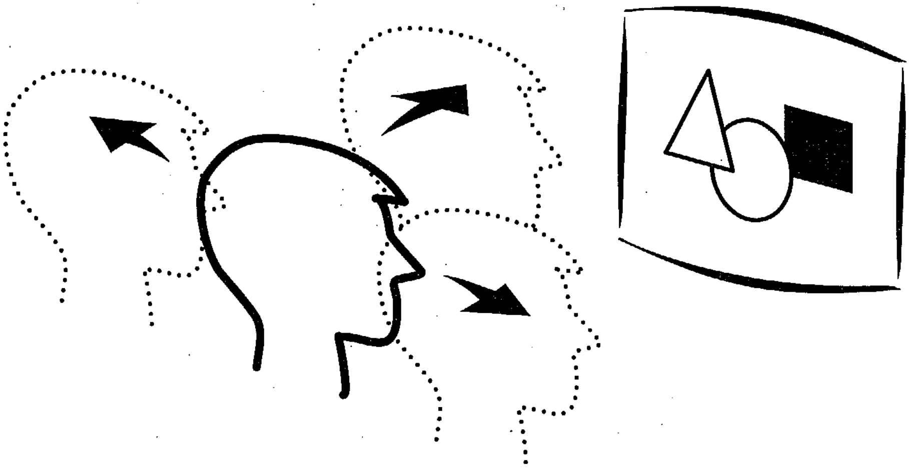
Another popular technique exploits the directionality of lenticular screens. By vertically alternating strips of stereoscopic images, the lenticular screen directs the corresponding view to each eye. We see these on "3D" greeting cards. At least one camera, made by Nimslow, was available to consumers. It had five lenses rather than two to create smoother parallax (technically making it a multiscopic camera, see below).

The stereoscopic movie craze became popular in the early 1950s when movie theaters began to compete against home television. These films use polarized glasses which separate the two simultaneously projected stereo views. Often there is "leakage" from one image to the other creating a ghosting effect, as well as other perceptual ambiguities. The result was usually a headache for the viewer.

Often these films tend to overemphasize the depth in their story, as if to make sure the audience gets the full "3D" effect. Broomsticks sweeping in and out, bats flying toward the camera, and sinister faces leaning in close were common in the stereoscopic horror film genre (one can use their imagination for the sex film genre).

A personal experience: I saw *Jaws 3D* at a local movie theater several years ago when it was first-run. It used polarized glasses, and the technical quality of the projection was relatively good. I hated the film, I simply thought it was bad. A few years later, I stumbled upon it on tv. I watched it. It was pretty good.

III. Looking Through A Window From Multiple Points Of View: Multiscopic Imaging



I define multiscopic imaging as representing more than one point of view for a given image. Stereoscopic imaging, by definition, represents two points of view, one for each eye, and is equivalent to looking through two peepholes. But in the real world, when we move our head a bit, we have available a range of points of view representing left and right, up and down, and in and out, of at least a few inches. This section outlines such forms of imaging.

Multiscopic imaging represents multiple points of view: lateral head movement while the body is more or less still. The result is local motion parallax (global motion parallax equals travel). This local motion parallax is a stronger depth cue than stereoscopic parallax since more than two points of view can be seen in a relatively short period of time.

Local motion parallax occurs when we turn our heads because our eyes are displaced from our neck's axis of rotation. It's a feature of evolution shared by humans and many members of the animal world. If our eyes were *on* the neck's axis of rotation (like a camera mounted on a tripod), there would be no lateral displacement when we turn our head and therefore no parallax.

Examples:

Multiscopic imaging is very rare; we have probably seen but a handful of examples. Here are some techniques.

Mirrors

A mirror is multiscopic by nature. When we view ourselves in a mirror, each eye is seeing a different point of view, so we see stereoscopy. But also when we move our head, we see correspondingly different points of view.

An early technique for achieving multiscopic images involved a giant half-silvered mirror, hidden from view, which reflected hidden images and props. Like normal mirrors, both parallax and accommodation were preserved, but since it was half-silvered, the reflected imagery appeared transparent, making it great for ghosts, but not much else.

Such giant half-silvered mirrors date back to the Phantasmagoria shows of the 18th century. Perhaps its most popular current use is for the dancing ghosts in the grand ballroom of Disney's Haunted Mansion. Similar examples, but where the 3D floating images were composed of 2D film and video screens, occurred at the last three Worlds' Fairs: the GM "Spirit Lodge" (EXPO '86, Vancouver), the Australian Pavilion (EXPO '88, Brisbane), and the Ginko, Gas, and Mitsui Toshiba Pavilions (EXPO '90, Osaka). Often, the audience is told that the glass in front of the stage is there to protect the allergy-sensitive from the smoke, and please, no flash photography (which would reveal the inner workings via spurious reflections).

A technique for producing small multiscopic images employs a flexible vibrating mirror to rapidly change focal lengths. This varifocal mirror reflects a video display whose image is in sync with the vibration, resulting in a relatively small volumetric display.[15] Since the image must be from a computer-generated 3D model, direct display of camera-originated images is not possible. Indeed, no camera exists.

Relief Projection

Another multiscopic technique is called relief projection, where an image is projected onto a screen whose shape physically matches the image itself. The most popular application of relief projection is "talking heads," where a mask is made of a person's face to be used as the projection screen. The person is filmed with their head totally motionless but their mouth and eyes moving. The film is projected onto the face-mask "screen", with careful alignment so that the eyes fall in the eye sockets, the mouth along the mouth line, etc. The illusion is very powerful, and the fact that the image of the eyes and mouth moves and the screen does not is barely noticable. The talking head in the Haunted Mansion at Disneyland uses this technique. A more advanced version, where the mask screen moves in sync with the image, was produced at MIT in 1980.[16]

I have produced room-sized relief projection by painting entire livingrooms white after filming them and projecting the original image back on the white-painted surfaces.[17] This was as art. The limit of relief projection, of course, is that the shape of the screen cannot easily change.

One method of making more flexible relief projection display is to rapidly spin a disc or corkscrew-shaped screen while projecting on it with synchronized lasers.[18] The result is a volumetric display (usually inside a clear housing for safety) whose size, detail and flicker rate relate to the computational horsepower and mechanics of the system. And like varifocal mirror displays, all the imagery must come from a 3D computer model rather than directly from a camera.

Holography

Holography achieves both parallax and accommodation, but is a filmic medium (with the significant exception of Dr. Steven Benton's most recent work at the MIT Media Lab). Being film-based has it implications. It cannot be transmitted live like video. Also, it is near-impossible for any kind of computer control and interactivity.

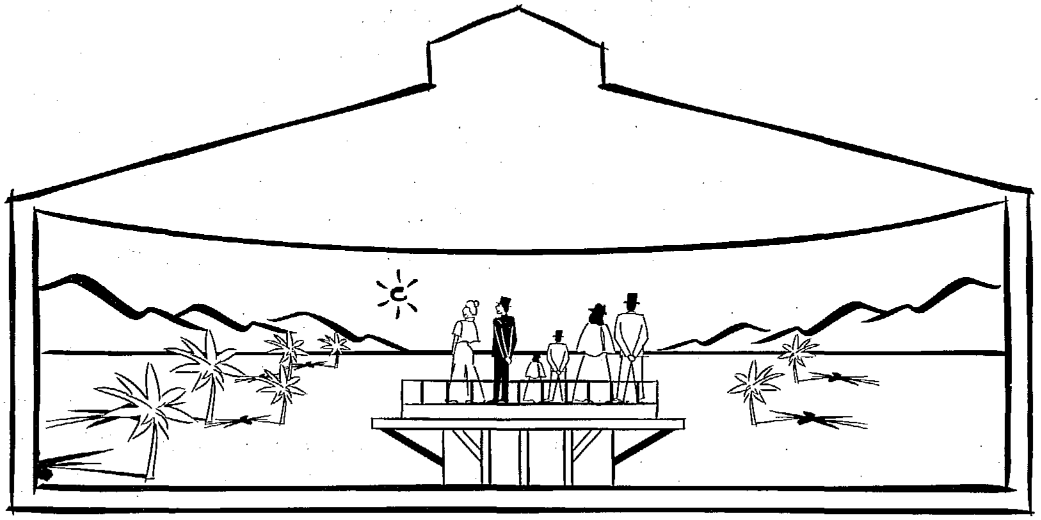
Another popular misconception about holography is its "projection." The holographic image can only be seen while viewing in the line of sight of the film: it may appear behind the film or in front of the film, but one must always be looking *through* the film. The concept of both the audience and the hologram being on the ground and a holographic image "projected" in the sky is simply inaccurate.

"Stereograms" (or "integrams") are holograms made from filmed or computer-generated material, where images are recorded from multiple points of view along straight or curved track. Like stereoscopic recording over time with a single camera, any motion counteracts the stereoscopy. Real holographic movies, though demonstrated, require massive amounts of holographic film and still offer very limited viewing.

Viewpoint-Dependent Imaging

Local motion parallax is possible with conventional display if it is driven by the user's head position. An example was produced at MIT, where an outdoor scene was shot laterally and mastered on videodisc.[19] The videodisc speed and direction was controlled by a single user wearing a position tracking device on his head. Because such a display is interactive in nature, it is limited to a single user. As the user sways back and forth, the video image changes accordingly. Though trivial with a virtual camera, recording with a realworld camera becomes increasingly more difficult as more than one dimension is shot (allowing the user to sway back and forth, in and out, and up and down simultaneously). And like other single camera applications for multiple points of view, time artifacts (motion) counteract the stereoscopic effect. Time must stand still.

IV. Looking Around: Panoramics



Panaromics is the ability to look around. An image is considered panoramic as it approaches framelessness, when the image is larger than the viewer's field of vision. And when this occurs, the feeling is that of *immersion*, of being inside, rather than outside looking in. Panoramic imagery allows freedom of angular motion, moving your head around.

Strictly speaking, a panoramic image is monoscopic because it represents the view from a single point in space, called the nodal point. The nodal point is the point in the camera where the angular coordinates (pan, tilt, and roll) are independent of the lateral coordinates (x,y, and z). At the nodal point there is no lateral displacement, and therefore no parallax. In motion control terms, panoramics allow the viewer control of the angular coordinates but not the lateral coordinates.

Exception:

Cubism was an explicit attempt to depict a scene from multiple points of view simultaneously. Front and side views of the same face in Picasso's cubist and post-cubist work are examples of depicting a scene from more than a single nodal point. Braque spoke of "inverting the pyramid of vision" of Renaissance perspective.[20]

History

The concept of painting enormous wrap-around images was first realized by Robert Barker, who patented it in Edinburgh in 1787. Barker's design, a custom-built cylindrical building with a dark stairway leading up to a central viewing platform, set the standard for the panoramas (later called cycloramas) to follow. His first panorama opened in London in 1792 and was an instant hit. From then until the birth of cinema in the 1890s, panoramas and cycloramas were the most popular art entertainment in the United States and Europe. They were often huge, 50 feet high and 400 feet in circumference, and often depicted scenes of high drama, such as the Battle of Waterloo or the Crucifixion.[21]

The first panoramic film system was patented only two years after the birth of projected cinema, by Raoul Brimoin-Sanson in 1897 in France. His Cineorama system used ten synchronized projectors projecting on a circular wall over 300 feet in circumference. The first giant rectangular screen was developed by the Lumiere Brothers for the Paris Exposition of 1900, which was 70 feet wide and 53 feet high. The first three screen projection system was developed in 1927 by Abel Gance for his epic film *Napoleon*, and the first largescale fisheye projection onto a dome was "Spacarium," in the 1962 Seattle World's Fair.[22] All of these systems have had modern descendants.

Exception:

Contrary to popular belief, Abel Gance's three screen "Polyvision" system which he developed for *Napoleon* was not intended for creating composite panoramic scenes as much as it was for displaying multiple images side-by-side.[23] Gance was creating in space what Eisenstein's "collisions" (his term for montage, or cuts), were creating in time.

Types of Perspectives

To understand panoramics, one must first understand a bit about perspective.

Rectilinear Perspective

Rectilinear perspective is shot on flat film and displayed on a flat surface. When viewing rectilinear perspective images off axis, the frame will appear trapezoidal, but straight lines will always appear straight. Practically every camera we have ever seen or used and practically every image we've ever seen or made, has been of rectilinear perspective.

Examples:

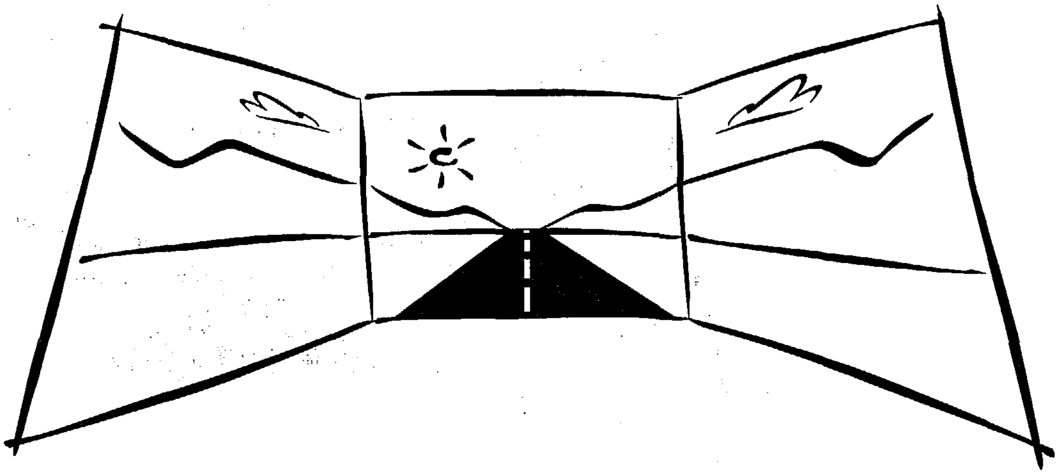
"Widescreen" is a designation for films shot with aspect ratios greater than 1.33 to 1 (1.33:1). Standard theatrical films have an aspect ratio of 1.85:1. The widescreen format popularized in the early 1950s by Cinemascope has an aspect ratio of 2.35:1 and often uses an "anamorphic lens" which "squeezes" the image horizontally by a factor of two.* Other widescreen formats include Panavision and Todd-AO 35 (also 35mm anamorphic but often release in 70mm non-anamorphic), Super Panavision and Todd-AO (70mm non-anamorphic), VistaVision (35mm run horizontally through the camera), and IMAX, the largest flat film (70mm run horizontally through the camera).[24]

The recent trend of video walls is noteworthy. Most video walls break a single video signal into the corresponding components for display over a matrix of video displays. The result is a larger image (usually with visible mullions, black gaps between images), but not greater overall resolution. Video walls are a good example of the importance of scale over resolution.

*** The term "anamorphic" technically refers to all lenses which "distort", like fisheyes for example, but in the context of widescreen, "anamorphic" is used to specifically refer to a horizontal compression. Also in the widescreen context, a non-anamorphic lens is referred to as a spherical lens, even though it has a rectilinear perspective. I will reserve the term "spherical" to only refer to lenses which produce spherical perspective.**

Perhaps the livliest argument surrounding high definition television standards is about aspect ratio. Most HDTV proponents agree that it should be wider than normal (4x3 or 1.33:1) television, but just exactly how much has spawned heated debates. The NHK standard for HDTV, the first major standard, originally had an aspect ratio of 5x3 (1.66:1). NHK researchers stated bluntly that they determined it to be optimal, though their research was acknowledged as "subjective".[25]

The Soviets developed an alternative to fixed aspect ratio cinema in the 1960s which they called "varioscopic cinematography." Both camera and projector were fitted with several fixed aspect ratio lenses ranging from 2.35:1 to .46:1 and could change between scenes.[26] This system is distinctively Russian in that it conforms composition to content, and is equally un-American in that it suggests not filling the entire screen all the time.

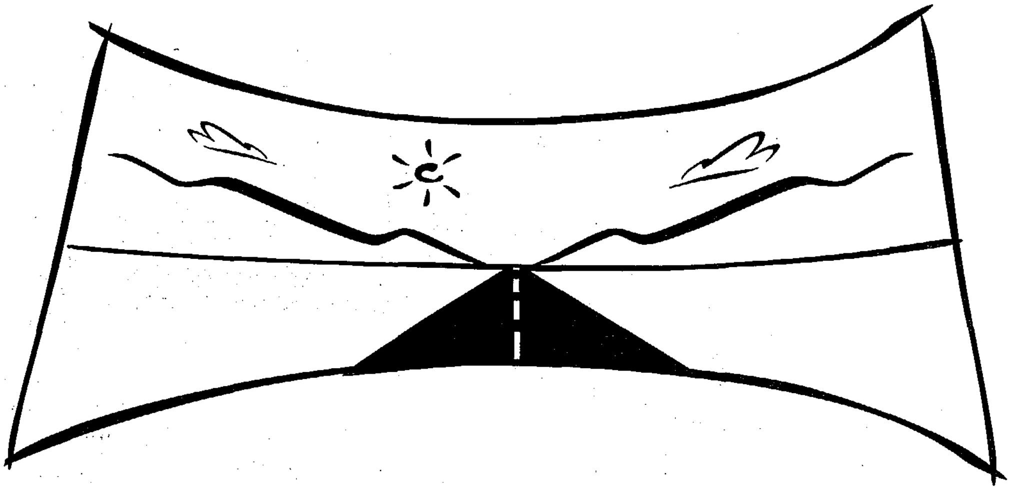


Because of their flat nature, all rectilinear perspective images must have less than a 180 degree angle of view, and therefore full panoramic construction requires multiple images. MIT's Aspen Moviemap project was shot with four 16mm cameras with just under 90 degree lenses, pointing front, back, left, and right. For these images to be viewed together properly, one must stand in the center of a four-walled projection space, otherwise trapezoidal distortion will occur. The four images, when laid flat, will exhibit discontinuities that can be corrected by "undistorting" them linearly with a computer for user-controlled panning on a flat display.[27]

Cylindrical Perspective

Cylindrical perspective is shot on cylindrically positioned film and displayed on a cylindrical surface. Unlike rectilinear perspective, only one cylindrical perspective image is required for a 360 degree panorama, and since it is a single image, there are no discontinuities. Cameras with rotating slits and lenses (such as the Widelux, Hulcher, Globus and Roundshot cameras) can shoot a single image over a relatively short period of time.

Optimal viewing is from the center of the cylinder, like being inside a giant lampshade. When the view is off-axis, straight horizontal lines will appear curved, while the vertical lines will appear unaffected. The distortion can be "dewarped" with a computer by nonlinear correction in one dimension and linear correction in the other dimension for user-controlled panning on a flat display.



Examples:

Disney's CircleVision, a direct descendent of the Brimoin-Sanson system, was shot with a pod of nine normal rectilinear 35mm cameras, each aimed upward at a mirror so that all cameras share a common nodal point. Instead of placing the projectors in the center of the theater, they were hidden in the mullions opposite each screen. An odd number of screens was required. Each of the nine screens, though cylindrical, were apparently flat enough not to exhibit significant distortion.

Like the other enhanced-format theaters, CircleVision emphasizes what it does best: showing sweeping vistas, dramatic landscapes, etc. The films tend to be mostly travelogues with names like "This is . . . .!"

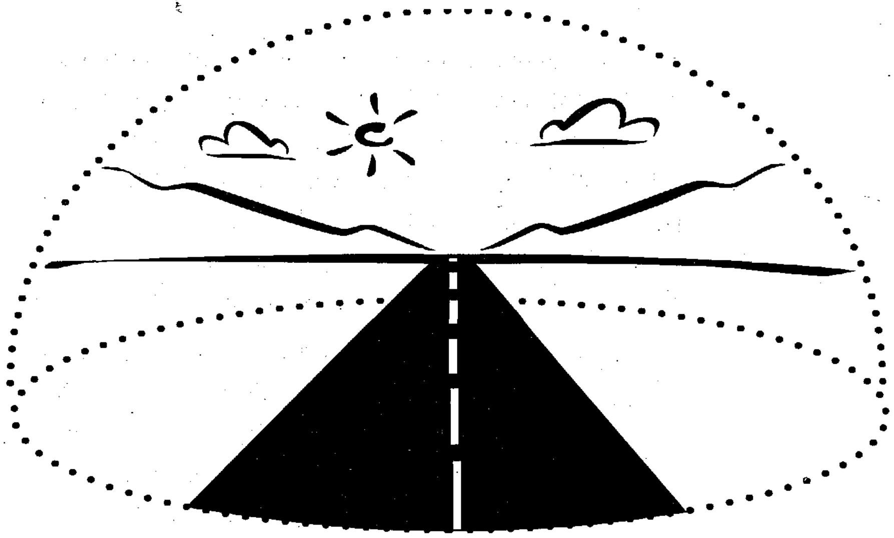
Cinerama, a descendent of Gance's Polyvision, was a three-screen system with an angle of view 55 by 146 degrees, approximating the angle of human sight. Also shot with standard 35mm cameras, Cinerama's distortion was more noticeable for several reasons. The screen was more curved per projector than CircleVision, so unless you were the lucky person sitting in the one proper seat, the image would appear curved. Also, Cinerama had no mullions, all three images combined to form one continuous image, in theory at least. The "line problem," as it was called, was always noticeable, and cinematographers were often forced to compose their shots to lessen this effect.[28]

Spherical Perspective

Spherical perspective is shot with special spherical optics (such as fisheye lenses) and displayed on a spherical surface. Optimal viewing is from the center of the sphere, like being inside a large dome or planetarium. When the viewer is off-axis, straight lines in both directions will appear curved.

Spherical recording is most often associated with fisheye lenses, but other such specialty lenses exist. The Peri-Appolar lens made by Volpi was used for the Aspen Moviemap. It produces a donut-shaped image representing a 360 degrees azimuth by ± 30 degree elevation, making it centered on the horizon rather than on the zenith when pointing upward. Shooting off of convex mirrors also produces spherical perspective (the Legg's pantyhose package is a favorite), but the camera will be visible in the middle of the frame.

Spherical perspective imaging captures the most in a single shot, but flat viewing results in severe distortion, as demonstrated by fisheye photos in magazines. The distortion can be "dewarped" with a computer by non-linear correction in both dimensions.



Examples:

Spherical perspective display, in the case of fisheye lenses, requires a dome. The popular Omnimax system (like IMAX, using 70mm film projected horizontally) premiered in 1973 in San Diego and projects on a dome tilted toward the audience at an angle of 25 degrees.[29] Sitting off axis results in distortions in both dimensions.

Substituting Interactivity for Wholeness

For each type of perspective, it is possible to store the entire panoramic image in such a way that the user may access a subset of it. The obvious advantages are that it eliminates the need for a 360 degree projection space and that it requires less display bandwidth.

Examples:

A reasonable method of viewing panoramic imagery is through a small window such as a video display as long as the user has control of the point of view, using a joystick, for example. Intel's DVI technology has such a method for "dewarping" and displaying imagery shot with a fisheye lens. The advantage is convenience. The loss is a spatial correspondence between the virtual space and the playback space.

A less reasonable method (but one that kept the author obsessed for several years) is where a projected image physically moves around the playback space in order to retain the spatial correspondence. When the camera panned left 90 degrees, the projected image also pans 90 degrees. The effect is exactly like using a flashlight in the dark.[30] Though such a "moving movie" requires neither the power nor the bandwidth to fill the entire playback space at once, it nevertheless requires a special playback space.*

Now imagine wearing the same system on your head. And instead of controlling pan, tilt, and roll with a joystick, you are controlling it by the position of your head. Wherever you look, the image will appear in front of your eyes. This is exactly what head-mounted displays (HMD's) do, and they can do it with properly accommodated, stereoscopic, wide angle optics.[31]

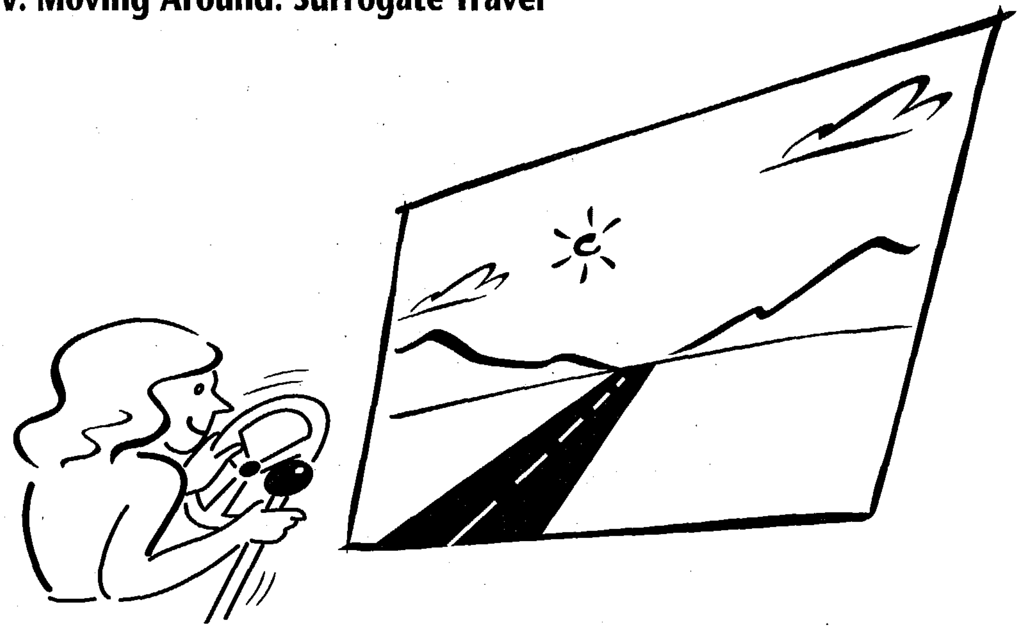
The first observation often made by users of HMD's is that there is an apparent simultaneity or "tight linkage" between head position and the system's visual output. It feels good. When there is not a tight linkage (lag, for example), the system feels "wrong." (Imagine turning your head in the real world and having the image lag.) When controlling a panoramic window interactively, tight linkage is desirable. Tight linkage, by definition, requires a relatively high degree of temporal resolution, independent of spatial resolution. To put it another way: apparent simultaneity requires fast iterative sampling.

Virtually all imagery shown in HMD's today are either computer-generated or from live telerobotic cameras. Realworld recording and storage for HMD's presents novel challenges. For example, shooting for both stereoscopy and panoramics has no simple solution, since two panoramic cameras separated for stereoscopy results in variable parallax as the "interactive small windows" rotate.

---

* Consider what happens when we fixate on an object and pan and tilt our head. A remarkable servo-control relationship occurs between our eye position muscles and our intention to keep our fixation. This "gyro-stabilization" effect can be easily seen by looking at your own eyes in a mirror and moving our head back and forth and up and down: your eyes stay fixed. Even more incredible is what happens when we rotate our head. Though muscles may compensate for rotation a bit, clearly when we rotate our head ninety degrees, the compensation turns from servo-muscle to cognitive. Readers may wish to experiment in front of their bathroom mirror.

V. Moving Around: Surrogate Travel



Surrogate travel, or "moviemaps," is imagery which allows the user to move laterally around a recorded or created place. Moving around any virtual space presents some problems not present when looking around a panoramic scene. Looking around need not be explicitly interactive: the entire view can be displayed and you may control what you see by moving your head. But moving around under your own control *must* be explicitly interactive: you must tell the system to change the lateral position. Thus, while an audience in a panoramic theater can all look in different directions, an audience in a "surrogate travel" theater will somehow have to come to grips with who's controlling the travel.

It is noteworthy that high-end "dark rides," such as in theme parks, rarely recreate a sense of place using any electronic or filmic media, but generally rely on traditional theatrical props and painted backdrops.

Control of surrogate travel doesn't require the same realworld metaphor as control of panoramics may have, like moving one's head. Most surrogate travel systems incorporate some sort of "remote vehicle" metaphor, with controls such as joysticks and speed levers, pushbuttons and turn indicators. An exception is the Exploratorium's Golden Gate exhibit, produced by Advanced Interaction Inc. and directed by the author, where the principle input device is a trackball, creating a tight linkage between one's hand and the speed and direction of the image. Another notable exception is the use of hand signals while wearing "data gloves" for moving through "virtual reality" worlds.

Other problems arise when surrogate travel is shot with real cameras, rather than generated from 3D databases. Though it is possible to record an entire panorama from a given point in a single instant, the only way to record surrogate travel in a single instant is with one camera at each location. For the Aspen Moviemap, such recording would have taken about 27,000 cameras, all shooting at the same instant.[32] The obvious alternative is to move a single camera from one location to another, but time artifacts from moving clouds, shadows, cars, and people are unavoidable.

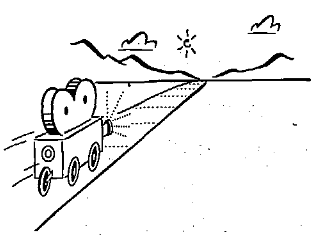
Another difference between panoramic recording and surrogate travel recording is "continuousness." Once a panoramic scene is recorded and stored as a continuous single image, the user may have continuous access to any portion of it, since it all exists in the single image. But surrogate travel requires lateral movement, recording single images at specific locations at regular intervals (like one frame every ten feet). Creating in-between images is a state-of-the-art problem, requiring heavy computing horsepower.

Surrogate travel material derived from real world recording is currently stored as a series of 2D images, hooked up to a fast access media system, such as optical videodiscs.

One-Dimensional Movement

One-dimensional surrogate travel is along one particular path. The user may go forward and backward, at any speed, but cannot stray from this path.

Distance-Dependent Images



In order to give the user a predictable sense of speed control, realworld images along the route are best shot at regular spatial intervals. Both film and video motion picture cameras are time-triggered instruments, recording one frame every 1/24 or 1/30 of a second. Distance triggering is desirable for surrogate travel. If the camera tracking speed can be held constant, then time triggering is equivalent. Otherwise, explicit triggering is necessary, such as from an odometer, or an external fifth wheel (as with the Aspen Moviemap).

The triggering distance affects visual continuity on the one hand and frame "real estate" on the other. The more images, the smoother the apparent movement, but the more storage space required. Smoothness is partially related to angle of view of the camera, height and distance to the nearest objects, and camera stability.

Image Stabilization

Image stability is also a realworld problem, not relevant for virtual cameras or for model cameras on motion control systems. Instability results from any variance of the lateral path or the angular position of the camera during shooting and falls into two classes: high frequency and low frequency. High frequency instabilities such as vibrations will produce blur or smear and affect individual frames. They can be minimized by using a wide angle lens, a short exposure time, and avoiding close or fast-moving objects.

Low frequency instabilities will produce a "wobble" from frame to frame. Since moviemaps are often sampled at rates less than normal motion pictures (one frame per second may be a nominal recording speed, for example), such instabilities are exaggerated. Consequently closed-loop gyroscopic stabilizers (such as Wescams, Gyrospheres or Tyler "Sea Mounts") perform better then either passive gyroscopic stabilizers (such as some helicopter mounts) or passive inertial stabilizers (such as Steadicams). In-camera and in-lens stabilizers (such as the 1962 "Dynalens", Arriflex's Image Stabilizer and Schwen's "Gyrozoom") can only correct for pan and tilt but not for rotation.

"1.1" Dimensional Movement

Moving along a path with occasional choice points is a far cry from being able to "travel anywhere." One may call this class of surrogate travel "1.1 dimensional" because only some of the points along the path have a two dimensional choice and most have a one dimensional choice.

Example:

If each block in Aspen averaged 300 feet between intersections, and if the street was shot one frame every ten feet, then there are 30 frames shot per block. Only one of those frames represent a node having the two dimensional choices of "forward, backward, left, or right" while the remaining 29 frames have only the one dimensional choice of "forward or backward." The "average dimension" is therefore [(29x1)+(1x2)]/30 or 1.03. In this case, calling the Aspen Moviemap "1.1 D" is being generous.

Match-Cuts

Suppose several paths were recorded and stored that occassionally intersect. If at the intersection points or nodes, the paths share an identical frame, then it is possible to "match-cut" from one path to the other invisibly. This is presumably the goal of surrogate travel: seamless user-control of movement and direction, like navigating through the real world.

The better the match-cuts between two intersecting routes, the greater the sense of seamlessness. Several factors contribute to matching cuts. First, the camera has to be in the same position and pointing in the same direction for both routes as it passes through the node. One may use lines on the street or trails and one may use compass coordinates, but there is no easy way to do this in the real world. (A field camera that knows where it is would be extremely helpful for surrogate travel.) Also, time artifacts are inevitable, since the matching images must be shot at different times. Lighting and shadow discrepancies can be minimized by shooting during a narrow window of time, like from 10 am to 2 pm, or shooting on cloudy days. For 3D database recording, as well as for motion control model shooting, these problems don't exist.

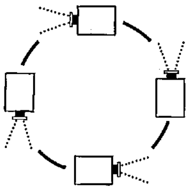
Examples:

Such match-cuts, cutting between the straight movement down a street and the turning movement through each intersection, formed the basis of the Aspen Moviemap. Paths with intersecting nodes, the "classic moviemap" style, also include the Bank Street and Sarnoff Lab "Polenque" project, and the Exploratorium's "Golden Gate" exhibit.

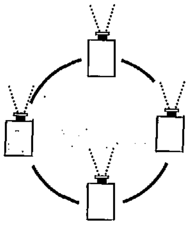
Exception:

The Paris Videoplan is a street-view moviemap of the Madeleine district of Paris produced by Logovision for the RATP (the Paris Metro) and directed by the author. Rather than filming turns, we filmed a mime standing in each intersection and pointing to the various options. We could then cut from the pointing mime to a view of the direction she was pointing. The concept was to replace perceptual continuity with cinematic continuity.

Camera Angle



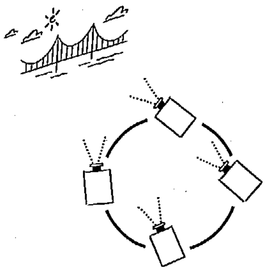
Since panoramic recording is often impractical, the camera's angular position becomes an issue, since a lens less than 360 degrees must be explicitly pointed. The simplest technique is to fix the camera angle to the lateral direction of motion, either pointing straight ahead or pointing sideways. At each node, each possible turn must be recorded for match cuts. MIT's Aspen Moviemap was shot in this fashion.[33]



Another more complicated way to point the camera is in an absolute direction independent of lateral position. For example, a camera could always point north regardless of whether it is facing forward, sideways, or backward with respect to lateral movement. One advantage is that shooting turns is not required since the camera always points in the same direction.

Example:

Most of Aspen Moviemap was shot with the principal camera facing forward, in the direction of travel. More complicated alternatives were investigated in Aspen where the camera anticipated the turn by panning in its direction before the vehicle.



An even more complex way to point the camera is at an absolute location such as tracking a central object. Absolute pointing requires independent control of the angular position as well as the lateral position of the camera. Like absolute position, the payoff is that separate turn sequences are not necessary since the camera always points in the same direction at any given point.

Example:

The Golden Gate videodisc exhibit was shot with the camera always pointing at an absolute location: the center of the Golden Gate Bridge. The grid was 10 x 10 miles in one mile increments, 1,000 feet above sea level, centered on the Golden Gate Bridge. North/South routes match-cut with east/west routes since the camera was always pointing in the same direction.

2-D and 3-D Movement

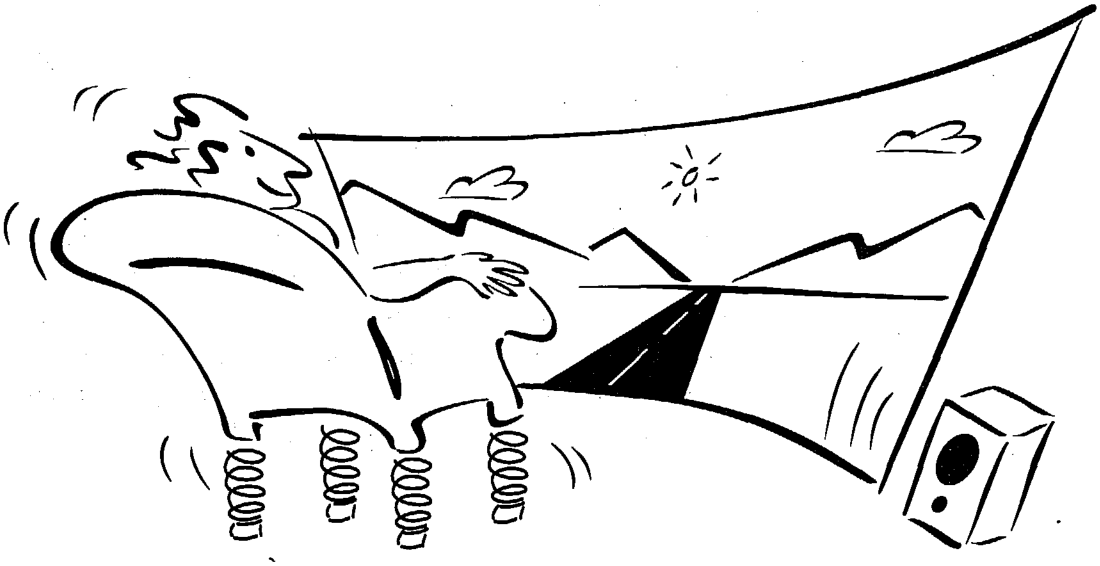
Recording and storing two-dimensional grids and three-dimensional lattices, where the user has freedom of movement, is problematic because the numbers grow quickly. Consider that a 15x20' space with a 10' high ceiling requires 3,000 frames if shot at intervals of one foot, and over 5 million frames if shot at intervals of one inch!

And we haven't discussed where the camera is pointing. For angular freedom, each of these shots would have to be a panoramic image. If a panoramic image requires 4 frames in order for a framed portion to have the resolution of a normal frame, we now need over 18 million frames.

In the future, the very idea of discrete frame storage will be obsolete. Computers will store information in spatial databases based on whatever data was collected (and will interpolate what is missing). It has been demonstrated that significant bandwidth compression occurs when the visual data is stored as a single computer model rather than as many, highly redundant, frames.[34] But data will still need to be collected, and visual data will be collected with cameras. The recording problem will never go away.

VI. Realtime Imaging: Motion

Realtime imaging is the process of recording and displaying *temporal* sensory information indistinguishable from unmediated reality.



Dynamic Visual Cues

Frame Rate

At least 15 updates per second are necessary for motion to appear on a screen. The upper level is arguable. Modern American film runs at 24 frames per second (fps), American video updates at 60 fps, and some argue that 80 or 90 fps may be necessary.[35]

Apparently, part of our association with the "film look" is film's *lack of* a sufficient frame rate. An often-discussed example occurred in 1976 when ABC produced their first (and last) made-for-tv movie shot in video rather than in film, "Victory at Entebbe," where they were rumored to have received letters complaining that the movie had a "soap-opera" look. Similarly, when video is defluttered (every other field removed reducing the effective frame rate from 60 to 30 fps), the result takes on this film look.[36]

Temporal undersampling may also be used to conserve frame "real estate", for example in the case of videodiscs. Traditionally it was far easier to keep a projector or videotape deck running at a constant speed and repeat static images than stop and start the mechanisms to conserve film or tape. No longer the case, the question of how many frames are necessary to convey the message has surfaced. Various experiments have been conducted with such compression[37] but more are clearly needed.

A related issue to frame rate is that of realtime playback. "Realtime" has an obvious intuitive definition in motion picture display: when the image appears "not too slow and not too fast," that is, when there is a one-to-one correspondence between record time and playback time. (If the reader has made it this far it should be clear why the term "realspace" is equally intuitive.)

Another related issue to frame rate is motion blur, the amount of time sampled for each frame. In the mechanics of film, motion blur is related to shutter speed, a function of the frame rate and the angle of the opening of the shutter (which may vary from 10 degrees to beyond 180 degrees). Blur is generally considered desirable when viewing realtime motion, and undesirable when scrutinizing motion sequences at slower than realtime speed, as in the case of sports. Thus, shooting for linear cinema optimizes blur while shooting for browsabililty optimizes crispness.

Example:

The Showscan film format records and plays at 60 fps, a rate its inventor Douglas Trumbull found optimal.[38] This high frame rate combined with its large gauge 70mm film format and high quality six track sound produces a relatively high degree of presence. Ironically, the image has a "video look," since its frame rate is the same as video.

Just as stereoscopic movies emphasize depth, every Showscan film I've ever seen emphasizes speed: race cars, airplanes, skiing, rollercoasters, etc., etc.

Exceptions:

Realtime playback is independent of frame rate. Intentionally undersampled movies, at rates from 1 - 15 fps, have been trendy in rock videos. Realtime undersampled movies may even have irregular frame rates, as was the case with "Max Headroom."

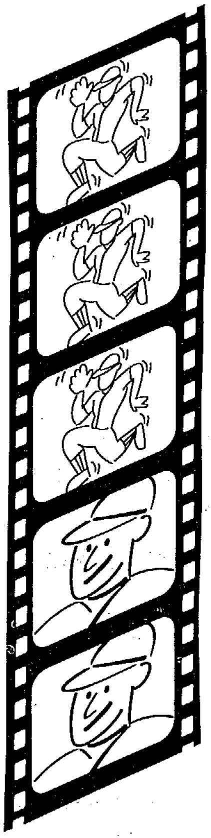
A classic example of temporally undersampled cinema is Chris Marker's *La Jette*, a 1964 French film. It is a dreamlike story shot entirely as a series of slideshow-style stillframes, with a single short exception. The exception, well into the film, is subtle, unexpected, and powerful.

Temporal Continuity

Temporal continuity, or seamlessness, is the opposite of cuts, or montage. The real world always exhibits perceptual continuity, regardless if it is seen looking out from a train or from a race car or sitting still. There are no cuts in the real world. There certainly may be cuts in our dream world and our cognitive world, but imagine, right now, looking up and *really* being somewhere else. Believing that you really are somewhere else (as opposed to imagining) is the dictionary definition of psychosis.

Cinema, on the other hand, consists of adjacent frames that either are temporally continuous (those within a shot) or not (those between shots, the "cuts"). Cinema is a counterpoint between "respect for spatial unity"[39] and its "first and foremost" characteristic, montage.[40] Though a great deal of research has been produced about temporal continuity and about montage, the distinction between the two surprisingly has been untouched.[41]

Examples:



"Classic" surrogate travel such as MIT's Aspen Moviemap relied on seamlessness between moving down the streets and turning from one street to another. It was precisely this seamlessness that gave it its credibility, its "realspace-ness." All flight simulators and most race car video games also rely on mimicking the temporally continuous nature of the real world.

Alfred Hitchcock's film *Rope* is a noteworthy cinematic example. It has virtually no cuts (in fact there are three). All action takes place in a single room and the camera is choreographed to gently move like an invisible voyeur. The camera moves were so well scripted that film reloading occurred as the camera brushed so close to something, clothing for example, that the momentary blackness could be used for the "match-cut."

Other noteworthy examples of temporal continuity are registered images, such as "then and now" pairs, timelapse in general, and different but similar objects such as faces registered by eyes. When these sequences are viewed over time, the better the registration, the more continuity, and the more successful they appear to be.

Exceptions:

All cuts in all films are violations of temporal continuity. The ability to go from Paris to Rome in 1/24 of a second is what makes cinema uniquely different from the real world. It has been argued that these violations, or "collisions,"[42] are where the art lies.

Perhaps the only everyday experience of perceptual discontinuities we have in the real world is waking up from a dream.

Dynamic Non-Visual Cues

Audio

Audio in synchronization with image is part of our association with cinema's ability to convey presence, and audio has its own resolution specifications. Of particular relevance to this survey is the spatialization of sound. Sound can be spatialized one of two ways: by using multiple speakers each positioned in the point of origin of the sound source or by using binaural sound.

Multiple channel audio has the advantage of accommodating groups of people moving around. In addition to the storage problems associated with independent channels, it is limited to stationary sounds, unless the speakers physically move.

Binaural audio retains the spatiality of the sound by recording or synthesizing the amplitude and phase differences from its sources to each ear. Binaural audio has two drawbacks: it requires headphones and the head cannot move unless extensive convolution to the audio is performed based on realtime input of head position.[43]

Inertial Motion

In addition to visual and auditory cues, we receive temporal cues by how we *feel*. This feeling of motion is based primarily in the vestibular system in the inner ear and is sensitive to linear and angular acceleration.[44] This feeling is most evident when we experience air flight (particularly when we are not looking out a window).

The most popular attraction today at Disneyland is Lucasfilm's "Star Tours," which exploits this parallel sensory channel. The forty-seat "spaceship" theater, unknown to the audience, is mounted on a large hydraulic system that can move the entire theater several feet in any direction with significant force. The motion is in sync with the picture. When the image tilts down, the theater tilts down, and so on. Such motion platforms were originally built for the flight simulator industry.

Even with such a motion platform, the effect is limited. It can only move in one direction for a short distance, limiting the duration and intensity of an acceleration. Furthermore, once it's at the end of its excursion, it has no place to go but back. The "Star Tours" story was carefully constructed to always go "back" after it went "forth," detectable to anyone able to pay such attention under the circumstances.

Force Feedback

Force feedback is the ability to actually *feel* a virtual object inside the image by virtually or remotely touching it. Virtual touching requires immersion in a virtual environment: you must be "inside" to touch anything there. Though several suggestions have been proposed, implementations require extensive mechanics.

Remote touching, receiving force feedback though the user is not visibly immersed inside the virtual environment, is a less formidable task. Such special devices have been built for testing. A force-feedback joystick has been used to simulate textures.[45] Similarly, a hand grip made of a four inch bar with three computer-actuated springs on each end can simulate angular and lateral force, and has been used successfully to augment visual display for spatial tasks.[46] In both instances, what one does is with their hand while what one sees is elsewhere, on a display. Such small, potentially practical devices are currently undergoing lively research.

Afterword

Each element of realspace imaging can be respected or violated. Either an image is orthoscopic, stereoscopic, or panoramic, or it's not. Sometimes violations of these elements are by default: it's more convenient to carry around "non-orthoscopic" images of your family in your wallet, stereoscopic cameras are expensive, and panoramic movies require special theaters.

But sometimes violations of these elements are intentional: a cut in a film, slow frame rate in sync with the music of a rock video, a simple line drawing rather than a high resolution image, silence rather than sound.

Trying to respect all the elements of realspace imaging is ultimately a losing battle. Giving the user everything is rarely possible. There is never enough bandwidth. There will always be artifacts.

The trick is to provide a *sense* of everything without actually giving everything. The question, then, is how do we choose what is most important? What is most important is always context-dependent.

This report was an attempt to lay out the choices. Choose wisely. *That* is where the art lies.

References

1. Mast, G., **A Short History of the Movies,** *Indianapolis: Bobbs-Merrill, 1976*
2. Minsky, M., **"Telepresence";** *Omni, Volume 2 (9), June 1980*
3. Papert, S., **Mindstorms: Children, Computers, and Powerful Ideas,** *New York: Basic Books Inc., 1980*
4. Kruger, M., **Artificial Reality,** *N.Y.: Addison Wesley 1982*
5. Fisher, S. S. et al., **"Virtual Environment Display System";** *ACM, Workshop on Interactive 3D Graphics, 1986*
6. Gibson, W., **Neuromancer,** *New York: Ace Science Fiction Books, 1984*
7. Sutherland, I., **"A Head-Mounted Three Dimensional Display";** *FJCC Vol. 33, 1968*
8. Wall Street Journal, *23 January 1990, front page*
9. Hochberg, J. and Gellman, I., **"Feature Saliency, 'Mental Rotation' Times and the Integration of Successive Views";** Memory and Cognition, No. 5, 1977
10. Schreiber, W. F., **"Psychophysics and the Improvement of Television Image Quality";** *SMPTE Journal, August, 1984*
11. Watts, A. W., **The Way of Zen,** *New York: Vintage Books, 1957*
12. Lipton, L., **Foundations of the Stereoscopic Cinema,** *New York: Van Nostrand Reinhold Company, 1982*
13. Bove, V. M., **Synthetic Movies Derived from Multi-Dimensional Image Sensors,** *Cambridge, MA: Ph.D dissertation, Media Laboratory, M.I.T., 1989*
14. Lipton, L., **Foundations of the Stereoscopic Cinema,** *New York: Van Nostrand Reinhold Company, 1982*
15. Design News, **"3D Display Without Glasses",** *p. 13, December 1988*
16. Negroponte, N., **"Media Room";** *SID 22, no. 2, 1981*
17. Reveaux, A., **"Displacing Time and Image";** *Artweek, June 30, 1984*
18. Williams, R. D. and Garcia, F., **"Volume Visualization Displays",** *Information Displays 5, no. 4, pp. 8-10, April 1989*
19. Fisher, S. S., **"Viewpoint Dependent Imaging: An Interactive Stereoscopic Display";** *SPIE, Volume 367, Processing and Display of Three-Dimensional Data, 1982*
20. Museum of Modern Art, *signage for the Braque and Picasso Show, New York 1988.*
21. Lawson, T., **"Time Bandits, Space Vampires";** *Art Forum, January 1988*
22. Katz, D., **"A Widescreen Chronology";** *The Velvet Light Trap Review of Cinema, No. 21, Summer 1985*
23. Bazin, A., **"The End of Montage";** *Cahiers du Cinema, No. 31, January 1954*

24 American Cinematographer Manual (F. H. Detmers, ed.), *Hollywood: The ASC Press, 1986*

25. Yuyama, I., **"Fundamental Requirements for High Definition Television Systems: Large- Screen Effects";** *NHK Technical Monograph No. 32, June 1982*
26. Wysotsky, M. Z., **Wide Screen Cinema and Stereophonic Sound,** *NY: Hastings House, 1971*
27. Yelick, S., **Anamorphic Image Processing,** *Cambrige, MA: B. S., thesis, Architecture Machine Group, M. I. T., 1980*
28. Katz, D., **"A Widescreen Chronology";** *The Velvet Light Trap Review of Cinema, No. 21, Summer 1985*
29. Chisholm, B., **"Widescreen Technologies";** *The Velvet Light Trap Review of Cinema, No. 21, Summer 1985*
30. Naimark, M., **"Spatial Correspondence in Motion Picture Display,** *SPIE. Volume 462, Optics and Entertainment, 1984*
31. Fisher, S. S. et al., **"Virtual Environment Display System";** *ACM, Workshop on Interactive 3D Graphic*
32. Mohl, R., **Cognitive Space in the Interactive Moviemap: An Investigation of Spatial Learning in Virtual Environments;** *PhD dissertation, Education and Media Technology, M.I.T., 1981*
33. Mohl, R., **Cognitive Space in the Interactive Moviemap: An Investigation of Spatial Learning in Virtual Environments;** *PhD dissertation, Education and Media Technology, M.I.T., 1981*
34. Bove, V. M., **Synthetic Movies Derived from Multi-Dimensional Image Sensors,** *Cambridge, MA: Ph.D dissertation, Media Laboratory, M.I.T., 19893*
35. Schreiber, W. F., **"Psychophysics and the Improvement of Television Image Quality";** *SMPTE Journal, August, 1984*
36. Naimark, M., **"Videodisc Production of the Visual Almanac";** *Technical Report #14, The Multimedia Lab, Apple Computer, c. 1988*
37. Naimark, M., **"Videodisc Production of the Visual Almanac";** *Technical Report #14, The Multimedia Lab, Apple Computer, c. 1988*
38. Turner, G. E., **"Showscan: Doug Trumbull's new 70mm Format";** *American Cinematographer, Volume 65 (8), 1984*
39. Bazin, A., **What is Cinema?,** *Volume 1, Berkeley: University of California Press, 1967*

40. Eisenstein, S., **Film Form,** *New York: Harcourt Brace Jovanovich, Inc., 1949*
41. Hochberg, J. and Brooks, V. **"The Perception of Motion Pictures"** *Handbook of Perception,* Volume X, New York: Academic Press, 1978.
42. Eisenstein, S., **Film Form,** *New York: Harcourt Brace Jovanovich, Inc., 1949*
43. Wenzel, Em. et al, **"A Virtual Display System for Conveying Three-Dimensional Acoustic Information";** *Proceedings of the Human Factors Society, 32nd Annual Meeting, 1988*
44. Deyo, R. and Ingebretsen, D., **"Notes on Real-Time Vehicle Simulation";** *ACM Siggraph, Course Notes 29, Implementing and Interacting with Real-Time Microworlds, 1989*
45. Minsky, M., **"Texture Identification Experiments";** *ACM Siggraph, Course Notes 29, Implementing and Interacting with Real-Time Microworlds, 1989*
46. Ming, O. et al., **"Force Display Performs Better than Visual Display in a Simple 6-D Docking Task";** *ACM Siggraph, Course Notes 29, Implementing and Interacting with Real-Time Microworlds, 1989*

Biography

Michael Naimark is an independent who has managed to straddle between the research community and the arts community for the past twelve years.

He was instrumental in making the first videodiscs for M.I.T., Atari, Siggraph, the Paris Metro, National Geographic, Lucasfilm, and the Apple Multimedia Lab. His specialty is cinematography, particularly surrogate travel. He has shot Aspen from the street, Paris from the sidewalk, and San Francisco from the air.

His artwork has been exhibited at the New York Avant Garde Festival, M.I.T.'s Center for Advanced Visual Studies, the San Francisco Museum of Modern Art, and the Exploratorium. He has served as art faculty at San Francisco State University, California Institute of the Arts, and the San Francisco Art Institute.

Currently, he is preparing a moviemap of Karlsruhe, Germany for a new arts and media center. He is also on the editorial board for the first scholarly journal on virtual reality, PRESENCE: Teleoperators and Virtual Environments published by MIT Press.

Michael received a B.S. in Cybernetic Systems from the University of Michigan in 1974 and an M.S. in Visual Studies and Environmental Art from M.I.T. in 1979.

Report Credits

Author:
Michael Naimark

Production:
Apple Multimedia Lab
  Kristina Hooper Woolsey
  Catherine Boyle
  Kristee Rosendahl
  Cindy Rink, Editor
  Art Jagonosi, Layout & Graphics

Pre-Production:
TechArt