michael naimark

| home **|** projects **| writings |** biography **|**

A 3D Moviemap and a 3D Panorama

Michael Naimark
Interval Research Corporation
Palo Alto, California

ABSTRACT

Two immersive virtual environments produced as art installations investigate "sense of place" in different but complimentary ways. One is a stereoscopic moviemap, the other a stereoscopic panorama. Moviemaps are interactive systems which allow "travel" along pre-recorded routes with some control over speed and direction. Panoramas are 360 degree visual representations dating back to the late 18th century but which have recently experienced renewed interest due to "virtual reality" systems. Moviemaps allow "moving around" while panoramas allow "looking around," but to date there has been little or no attempt to produce either in stereo from camera-based material.

"See Banff!" (1993-4) is a stereoscopic moviemap about landscape, tourism, and growth in the Canadian Rocky Mountains. It was filmed with twin 16mm cameras and displayed as a single-user experience housed in a cabinet resembling a century-old kinetoscope, with a crank on the side for "moving through" the material.

"BE NOW HERE (Welcome to the Neighborhood)" (1995-6) is a stereoscopic panorama filmed in public gathering places around the world, based upon the UNESCO World Heritage "In Danger" list. It was filmed with twin 35mm motion picture cameras on a rotating tripod and displayed using a synchronized rotating floor.

Keywords: moviemaps, panoramas, immersive virtual environments, art installations

1. CAMERA-BASED IMMERSION

"Immersion," in the context of media and virtual environments, is often defined as the feeling of "presence" or "being there," of being "inside" rather than "outside looking in." Various attempts have been made to taxonomize the elements required for presence, [1,2,3,4]

but at the very least studies suggest that visual presence is directly related to field of view (FOV). [5,6] Stereopsis and orthoscopy (i.e., where the viewing FOV matches the recording FOV, thus maintaining proper scale) often enhance immersion. [7] A variety of special-venue film formats have been developed to deliver immersive experiences. [8] These experiences are camera-based and represent the physical world, but are linear and don't allow any interaction or navigability. Computer-based immersive virtual environments do allow interaction and navigability, but are restricted to whatever can be made into 3D computer models. Since such computer models are built from scratch, they often represent imaginary or fantasy environments. Making computer models of actual places has proven to be non-trivial, and even today's best models display unwanted artifacts.

The work described here represents something between filmic and computer graphic immersion. It is camera-based and the imagery is of the physical world (in the documentary or ethnographic tradition), but it also has elements of interactivity and navigability, be they constrained.

2. MOVIEMAPS

Moviemaps allow virtual travel through pre-recorded spaces. Routes are pre-determined and filmed with a stop-frame camera triggered by distance rather than by time (typically done via an encoder on a wheel). Distance-triggering maintains constant speeds during playback at constant frame rates, which is often not practical or possible during production with a conventional (time-triggered) movie camera. The result, in a very real sense, is the transfer of speed control from the producer to the end-user, who has control over frame rate through an input device like a joystick or trackball.

In addition to speed control, limited control of direction is possible by filming registered turns at intersections. By match-cutting between a straight sequence and a turn sequence, the user can "move" from one route to another. Care must be taken to minimize visual discontinuities such as sun position and transient objects (e.g., cars and people). The goal is to make the cuts appear as seamless as possible.

2.1 Past Moviemaps

The first interactive moviemap was produced at MIT in the late 1970s of Aspen, Colorado**.** A gyroscopic stabilizer with 16mm stop-frame cameras was mounted on top of a camera car and a fifth wheel with an encoder triggered the cameras every 10 feet. Filming took place daily between 10 AM and 2 PM to minimize lighting discrepancies. The camera car carefully drove down the center of the street for registered match-cuts. In addition to the basic "travel" footage, panoramic camera experiments, thousands of still frames, audio, and data were collected. The playback system required several laserdisc players, a computer, and a touch screen display. Very wide-angle lenses were used for filming, and some attempts at orthoscopic playback were made. [9]

The author has since conceived and directed several moviemap productions, each with its own unique playback configuration. The "Paris VideoPlan" (1986) was commissioned by the RATP (Paris Metro) to map the Madeleine district of Paris from the point-of-view of walking down the sidewalk. It was filmed with a stop-frame 35mm camera mounted on an electric cart, filming one frame every 2 meters. An encoder was attached to one of the cart's axles. Rather than filming all the turn possibilities at each intersection, a mime was employed to stand in each intersection and simply point in the possible turn directions. The idea was to substitute the perceptual continuity of actual match-cuts with cinematic continuity. The playback system was built in a kiosk and exhibited in the Madeleine Metro Station.

The "Golden Gate Videodisc Exhibit" (1987) was produced for San Francisco's Exploratorium as an aerial moviemap over a 10 by 10 mile grid of the Bay Area. It was filmed with a gyro-stabilized 35mm motion picture camera on a helicopter, which flew at

a constant ground speed and altitude along one-mile grid lines determined by LORAN satellite navigation technology, effectively filming one frame every 30 feet. The camera was always pointed at the center of the Golden Gate Bridge, hence no turn sequences were necessary since the images always matched at each intersection regardless of travel direction. The playback system used a trackball to control both speed and direction, with the feel of "tight linkage" to the laserdiscs. The result was the sensation of moving smoothly over the Bay Area at speeds much faster than normal.

"VBK: A Moviemap of Karlsruhe" was commissioned by the Zentrum fur Kunst und Medientechnologie (ZKM). Karlsruhe, Germany, has a well-known tramway system, with over 100 kilometers of track snaking from the downtown pedestrian area out into the Black Forest. A 16mm stop-frame camera was mounted in front of a tram car and interfaced to the tram's odometer. Triggering was programmed to be at 2, 4, or 8 meter increments per frame depending on location. Filming on a track resulted in virtually perfect spatial registration. The playback system consisted of a pedestal with a throttle for speed control and 3 pushbuttons for choosing direction at intersections. The camera had a very wide-angle lens (85 degree horizontal FOV) and playback employed a 16 foot wide video projection. The input pedestal was strategically placed in front of the screen to achieve orthoscopically correct viewing, resulting in a strong sense of visual immersion.

2.2 The "See Banff!" Kinetoscope

But the immersive experience with the Karlsruhe Moviemap was monoscopic. One might argue that binocular disparity is not an important factor for landscape imagery (e.g., compared to infinity focus or motion disparity), but no one had yet made a stereoscopic moviemap.

In 1992, the author was working on field recording studies in the "Art and Virtual Environments" program at the Banff Centre for the Arts, and after several attempts at making computer models from camera-based imagery, reverted back to using moviemap production techniques, but this time in stereo.[10] The concept was to film scenic routes in and around the Banff region of the Canadian Rocky Mountains.

2.2.1 Camera Design and Production

The camera rig had to be small, portable, and rugged. As attractive as gyro stabilization may have been, it would have been much too heavy to take down mountain trails, on glaciers, and over narrow bridges. Without such stabilization, the stability of the imagery would be at the mercy of the terrain and, to some extent, the skill of the operator.

The basis of the camera rig was a 3-wheeled "super jogger" baby carriage, reinforced for extra sturdiness and modified to hold a tripod (see Figure 1). An encoder was installed on one of the rear wheels, with electronics for triggering the cameras from 1 frame every centimeter on up. A custom mount was built to hold two 16mm stop-frame cameras in parallel so that they could be released for film loading but would mount back in precisely the same position.

The cameras were fitted with the 85 degree horizontal FOV wide-angle lenses, with the intention of making a wide-angle orthostereoscopic display system.

The cameras were always triggered in sync. The stop-frame motors rotated at 1/8 second. The shutters were variable and modified to lock at 30 degrees, resulting in a shutter speed of 1/96 second, enough to freeze most motion if the rig moved at walking speed.

Figure 1. The "See Banff!" camera rig.

After much theoretical debate about optimal interocular distance between cameras, the minimum practical distance the cameras could be mounted apart was about 8 inches due to the size of the stop-frame motors. After reviewing several hundred turn-of-the-century landscape stereograms, it became clear that the exaggerated depth resulting from abnormally large interocular distances was the rule more than the exception, so no sleep was lost over our rig design.

The intent of production was to film a wide variety of unconnected routes without any intersections. The playback system would require speed control and route selection but not control over turns. This decision broadened the range of possible filming, since the notion of completion (i.e., covering all possibilities, filming turns, completing grids) wasn't necessary.

As it turned out, capturing the beauty of the landscape became overshadowed by the presence of tourists everywhere. Busload after busload appeared even in very remote areas. Dozens of people of all ages and cultures, clad in bright colors and toting cameras, wandered through the landscape. It became clear that, as an artwork, there was strength in counterpointing the beauty of the landscape with the actualities of tourism. The presence of tourists also created a lively foreground, giving the imagery a greater sense of 3D.

Filming took place during September 1993 in a wide range of locations. Stability ranged from finding smooth paved wheelchair trails to carrying the rig over rocky terrain. Frame rates were determined on-the-spot as a function of stability of the surface and distance of nearest objects, and ranged from 1 frame every centimeter to 1 frame every meter. Usually the cameras were pointed forward in the direction of movement but sometimes were pointed off to the side (and occassionally were stationary in a clock-driven timelapse mode). Over 120 scenes were filmed.

2.2.2 Playback System
Shortly after production, the film was transferred to videotape, edited, and transferred to 2

laserdiscs, one for each eye. A trackball-based interactive system was produced, with simple optics and mirrors arranged in a Wheatstone configuration for single-user stereoscopic viewing.

The first system approximated true orthoscopic viewing, with an 85 degree horizontal FOV. Such a wide FOV is considerably larger than sitting in the front row of the grandest of movie theaters, and the video resolution was extremely coarse spread out over such a large area. We backed off to a FOV closer to 60 degrees. It still looked very large and still appeared orthoscopically correct. (One is tempted to speculate that the human perceptual system acts like Saul Steinberg's "New Yorker's Map of the U.S., " that anything bigger or farther than that with which we are familiar appears so nonlinear that beyond some point it doesn't matter.)

The next step was to package it into a traveling exhibit. In roughing out a basic design - a podium-like box to house the hardware, an eyehood for single-user wide-angle orthostereo viewing, a one-dimensional input device - it became clear that a strikingly similar device had already been built. In April 1894, the Edison kinetoscope made its public debut. This was at a turbulent moment in the history of cinema, when the camera had already been actualized but projection had not. It was now exactly 100 years later, and the temptation to suggest an analogy was overwhelming (see Figure 2).

The final exhibit, called "See Banff!" (irony intentional), was built of walnut and brass in an authentic but exaggerated kinetoscope design, with a lever for selecting one of several scenes on exhibit and a crank for "moving through" the material. (History buffs will note that kinetoscopes never had cranks, mutoscopes did, in part because Edison was peddling another of his inventions, electricity and electric motors.) The crank was equipped with a force-feedback brake which freezes movement when the user reaches the boundaries of each scene, simulating film mechanics. (Perhaps not surprisingly, it feels "not right" when the brake is disabled.)

Partly for practical and partly for aesthetic reasons, a single laserdisc was used with field-sequential stereoscopic video and 30 Hz LCD shutters built into the optics. The flicker is noticeable, but after all, it *is* a kinetoscope.

3. PANORAMAS

The word "panorama" was coined in 1792 in London to describe the first of what became a popular form of public entertainment: a large

elevated cylindrical room entirely
covered by a 360 degree painting.11 From the beginning, a distinction was made between
displaying a panoramic image all at once (circular or stationary panoramas) and over time
(moving panoramas). This distinction has carried over to cinema as well.

3.1 "Moving Movies"
In 1977, the author began investigating what happens when a projected motion picture
image physically moves the same way as the original camera movement. If the angular
movement of the projector equals the angular movement of the camera, and if the FOVs
are equal, spatial correspondence is maintained, and the result appears as natural as
looking around a dark space with a flashlight.12 A simple demonstration was made using
a super-8 film camera and projector on a slowly rotating turntable. Later a more complex
system which recorded the pan and tilt axes was built. Finally a series of art installations
were realized using the simple turntable inside a space furnished to resemble a
livingroom, whose entire contents were spray-painted white after filming to become a 3D
relief projection screen for itself.13 This phenomena of motion picture projection
physically moving around a playback space was named "moving movies."

Moving movies have no lateral and only angular movements, which must be around the
camera's nodal point. As such, they are strictly 2D spatial representations. Computer-
based moving panoramas such as Apple Computer's Quicktime VR, Microsoft's
Surround Video, and Omniview's Photobubbles (to name a few) are similar insofar as
they rely on 2D representations from a single point of view, produced by tiling multiple
images or by using a single fisheye lens. Making any 2D visual representation from more
than one point of view results in distortion, by definition. (The essence of 2D
photographic representation and indeed, photorealism, is strictly from a single point of
view. Alternative forms like Cubism and David Hockney's photo-collages are fine
counter-examples.)

A 2D panorama represents a single point of view, but stereopsis requires at least two
points of view. Making stereoscopic panoramas with two 2D panoramic images is
problematic. For example, if two panoramic images are taken with stationary cameras
placed apart at an interocular distance, the disparity will vary as the user looks around
(and will become zero on the axis between the 2 cameras). One the other hand, if the
cameras move laterally during exposure to keep disparity constant, each camera's image
will no longer represent a single point of view, resulting in distortion (e.g., circles become
ovals, image appear fragmented, perspective becomes discontinuous).

3.2 "Be Now Here"
In December 1994, the author was invited to produce an art installation for the Center for
the Arts at the Yerba Buena Gardens in San Francisco to open in December 1995. The
installation proposed was called "Be Now Here (Welcome to the Neighborhood)." Just as
the Banff kinetoscope was an experiment in making a stereoscopic version of
interactively *moving* around, "Be Now Here" was to compliment it by making a
stereoscopic version of interactively *looking* around.

The concept was to assemble an experimental camera system to film stereoscopic
panoramas, then to go to public gathering places (*commons*) and film throughout the
course of a day from a single position. The experience would be analogous to standing in
a single place, with both eyes open, and being able to look around but not move from the
spot.

Site selection was based on the "In Danger" list issued by the UNESCO World Heritage
Centre in Paris.14 Of the 440 UNESCO-designated World Heritage Sites, 17 had been
further designated "In Danger." Of these, 4 are cities: Jerusalem, Dubrovnik (Croatia),
Timbuktu (Mali), and Angkor (Cambodia). With assistance of UNESCO, the plan would

be to visit each site, to work with local collaborators and to determine the most representative public commons and a single spot in which to set up the camera system. (Partly for art and partly for research reasons, going into interesting but fragile environments to make an statement about "place" seemed appropriate.)

3.2.1 Camera Design and Production

The camera design was based on 2 cameras (for stereo), 60 degree horizontal FOV lenses (for immersion), and a slowly rotating tripod (for panoramics), rotating once per minute (1 rpm). This is a compromise since it takes a minute to capture an entire 360 degree scene, but using multiple camera pairs to capture the entire scene at once was not practical. Sunlight variation was assumed to be negligible during the course of a minute, so using multiple images from the same scene for projection or panoramic tiling would have artifacts only from moving objects. A 1 rpm closed-loop crystal synchronized motor was mounted on a tripod.

The question of how to arrange the cameras with respect to the axis of rotation resulted in lively debate. Mimicking mammal head rotation suggested placing both cameras symmetrically in front of the axis of rotation. A colleague, John Woodfill, had a novel suggestion: rotate the camera pair around the nodal point of one of the cameras. Such a configuration would result in a perfect 2D panorama from one camera and would place all of the disparity difference in the other camera. As a vision researcher interested in the footage, he felt this could be useful. The "Woodfill Configuration" was deployed, with careful manual determination of one of the camera's nodal points. (One might speculate that looking at a stereoscopic panning movie where one eye sees no disparity and the other eye sees all the disparity would be noticeable, but one could counter-speculate that it's possible to sit on a rotating stool with one eye directly over the axis of rotation and conclude that there's nothing special about it.)

After much deliberation, it was decided to use 35mm motion picture film. The resolution would be 4 times greater than 16mm film and much greater than video, particularly with respect to dynamic range. It is also well-known that 35mm motion picture cameras are simple yet durable and time-tested, with less likelihood of failure than video in the field. Arriflex cameras with Zeiss lenses were selected. As with Banff, the size of the cameras made it difficult to obtain normal human interocular distance, so an exaggerated interoccular distance of 8 inches was used.

It was further decided to film at a frame rate of 60 frames per second (fps). Such footage could be transferred to video with each film frame corresponding to a single video field (half-frame). The result would have the best qualities of both film and video: it maintains the high dynamic range of film while having the motion smoothness of video (which updates at 60 fps). A sync box made by Cinematography Electronics was used to synchronize both cameras to a single controller, which allowed syncing phase as well as frame rate. The shutters were closed down to 30 degrees, resulting in an exposure time of 1/720 second, fast enough to freeze most everything.

Color negative film daylight-balanced with an ASA of 50 was used. Since all filming was to take place during daylight, the low ASA coupled with fast shutter speed wouldn't be a problem, with most filming possible at apertures between F4 and F11.

Using stereoscopic 35mm cameras with high quality lenses and low-speed film, running at 60 fps, and with synchronized shutter and rotational speeds would result in unrivaled fidelity. It would, at the very least, have 5 times the resolution of theatrical 35mm film (twice the spatial and 2.5 times the temporal resolution).

The complete camera rig, including cases, weighed 500 pounds but was built for travel (see Figure 3). All production took place during October 1995, including filming the Yerba Buena Gardens in San Francisco for counterpoint. A pro-DAT audio recorder

with a shotgun microphone was used to collect sounds at each site for later mixing into 4 channel rotating sound. Enough stock to film 5 panoramas (10 reels of 400' film) was taken to each site. Miraculously, production stayed on schedule and everything came out.[15]

Figure 3. The "Be Now Here" camera rig on location in Timbuktu.

3.2.2 Playback System
After production, the film was transferred to videotape, edited and mixed with the audio, then transferred to 2 laserdiscs. A simple input pedestal was made to allow site selection. Three scenes from each site were selected and aligned with each other such that perfectly registered time-of-day changes could be experienced within the same location.

Unlike Banff, Be Now Here would retain realtime motion and sacrifice browsability, enabling the user to control place and time but not speed. People movement would appear natural (not the case with See Banff) and coupled audio would be possible. This decision was based on the difference between footage of moving along long pathways (where browsability is desirable) and footage which, literally, goes around in circles.

A 12 by 16 foot highly reflective front projection screen would be used in conjunction with dual polarized video projection. The input pedestal would be strategically placed approximately 14 feet in front of the screen to entice the viewers to stand at the orthoscopically correct spot. The audience would wear inexpensive polarized glasses.

The obvious way to achieve spatially correspondent playback would be to rotate the projected image around the viewing space, another moving movie. But this proved impractical given the need for wide-angle projection. The solution: rather than rotate the projection around the static audience, to rotate the audience inside a static projection.

A 16 foot diameter rotating platform was used, rotating at 1 rpm in sync with the imagery (see Figure 4). The audience, limited to 10 at a time, stands on it (standing seems more desirable for ambient rather than narrative media). A black tent-like cylindrical structure

surrounded most of the viewing space to mask out the non-rotating world.

Figure 4. The "Be Now Here" installation.

The resulting effect is difficult to describe. Most viewers reported that after several seconds it felt like they were still and that the image was rotating around them. This effect is similar to the "moving train illusion," when a train sitting in the station pulls out and observers on the adjacent (non-moving) train believe their train is the moving one. Some viewers reported feeling the rotational force from the turntable, but most did not. (In a NASA design study on space colonies, it was determined that 1 rpm was the maximum rotational rate which would be undetectable by the general population. [16]) Though one might argue that physchophysical ambiguity exists between such audio-visual and vestibular cues, one could equally argue that a conventional panning image viewed in a conventional (non-rotating) movie theater produces the same degree of ambiguity between audio-visual and vestibular senses.

Most everyone reported feeling a strong visceral sense of place. And that's what the installation was about: conveying the feeling of presence by connecting our eyes and ears with the ground.

ACKNOWLEDGMENTS

REFERENCES

1. M. Naimark, "Elements of Realspace Imaging: a Proposed Taxonomy," SPIE Electronic Imaging Proceedings, vol. 1457, pp. 169-179, 1991.

2. T. B. Sheridan, "Musings on Telepresence and Virtual Presence," Presence, vol. 1, no. 1, pp. 120-126, 1992.

3. D. Zeltzer, "Autonomy, Interaction, and Presence," Presence, vol. 1, no. 1, pp. 127-132, 1992.

4. J. Steuer, "Defining Virtual Reality: Dimensions Determining Telepresence," Journal Of Communication, vol. 42, no. 4, 1992.

5. T. Hatada, H. Sakata, H. Kusaka, "Psychophysical Analysis of the 'Sensation of Reality' Induced by a Visual Wide-Field Display," SMPTE Journal, vol. 89, pp. 560-569, 1980.

6. C. Hendrix and W. Barfield, "Presence within Virtual Environments as a Function of Visual Display Parameters," Presence, vol. 5, no. 3, pp. 274-289, 1996.

7. E. M. Howlett, "Wide Angle Orthostereo," SPIE Electronic Imaging Proceedings, vol. 1256, 1990.

8. M. Naimark, "Expo '92 Seville," Presence, vol. 1, no. 3, pp. 364-369, 1992.

9. R. Mohl, Cognitive Space in the Interactive Movie Map: An Investigation of Spatial Learning in Virtual Environments, PhD dissertation, Education and Media Technology, M.I.T., 1981.

10. M. Naimark, "Field Recording Studies," Immersed in Technology: Art and Virtual Environments, M. A. Moser, ed., pp. 299-302, MIT Press, Cambridge, 1996.

11. A. Miller, "The Panorama, the Cinema, and the Emergence of the Spectacular," Wide Angle, vol. 18, no. 2, pp. 34-69, 1996.

12. M. Naimark, "Spatial Correspondence in Motion Picture Display," SPIE Proceedings, vol. 462, pp. 78-81, 1984.

13. M. Naimark, "Moving Movie," Aspen Center for the Arts, 1980; "Movie Room," Center for Advanced Visual Studies, M.I.T., 1980; "Displacements," San Francisco Museum of Modern Art, 1984.

14. See: http://www.unesco.org/whc/list.htm .

15. See: http://www.naimark.net/writing/trips/bnhtrip.html.

16. R. D. Johnson and C. Holbrow (ed.), Space Settlements: A Design Study, NASA SP-413, p. 22, 1977.

17. See: http://www.naimark.net/projects/banff.html .

18. See: http://www.naimark.net/projects/benowhere.html .