Movie-Maps: An Application of the Optical Videodisc to Computer Graphics

Andrew Lippman

Research Associate
Architecture Machine Group
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

ABSTRACT

An interactive, dynamic map has been built using videodisc technology to engage the user in a simulated "drive" through an unfamiliar space. The driver, or map reader, is presented with either sparsely sampled sequences of images taken by single frame cameras that replicate actual imagery from a space, or with computer synthesized replicas of those images. The reader may control the speed, route, angle of view and mode of presentation of this information and may thus tour the area. In addition, he may access spatially stored ancillary data stored in the buildings or in locales in the environment. This basic map is being enhanced to provide topographic views, and to incorporate optical and electronic image processing to provide a more responsive, visually complete representation of an environment.

Key Words: optical videodisc; computer generated imagery; anamorphic imagery; image processing; interactive systems.

Category Numbers: 3.80, 8.1, 8.2.

1.0 Introduction

In the past, research in computer graphics has centered on better, more rapid techniques for modeling and synthesizing realistic appearing images. Starting from calligraphic displays, in which the world was depicted as line segments, computer graphics grew to include the solids, textures, and colors possible with raster scan television systems, and developments in modeling have resulted in acceptably realistic and dynamic animation systems and flight simulators. Recent work has concentrated on the addition of synthesized texture (1), multiple illumination sources (2), and computer refraction (3), but pictures generated with these features must be done "off-line." The dynamics of flight simulator-type systems remains too costly for most applications.

Simultaneously, the television industry has begun to absorb the techniques of digital signal processing into their transmission systems, but with the different goal of real-time picture manipulation, and with a far less comprehensive set of modeling features. Framestore synchronizers, the broadcast industry analogue to raster scan computer displays, emphasize speed of operation, and implement image synchronization, smooth scaling and merging of pictures at the standard television rate of thirty images per second. These systems also use feedback to generate effects, but as a rule, accept real images as their data source, and provide real images as their output. Control rather than synthesis is stressed.

These two apparently disparate approaches to picture interaction and processing now have the capability to be merged by use of the optical videodisc. The manipulation expertise of the television industry can be combined with the modeling and interaction expertise of the computer graphics industry in systems that have the responsiveness and controllability of computer systems, but use the visually complete and detailed imagery of the television world. Further, these systems are available at costs that permit them to

---

The work reported herein has been supported by the Cybernetics Technology Division of the Defense Advanced Research Projects Agency, under Contract No. MDA-903-78-C-0039.

serve a great many applications. Unlike graphics systems which often are designed to meet certain performance goals at any price and are then developed into affordable products, the videodisc has progressed from an inexpensive movie player to an interfaced computer peripheral that provides random access to over 50,000 frames of television information.

A system that executes this merger, and at the same time explores its limits is the Movie-Map. At its simplest level, it may be regarded as a dynamic replacement for a topographic paper map: it can familiarize a user, or map reader, with a spatial environment. This familiarization, however, is accomplished by quite different means: the "map reader" explores space by participating in a simulated drive through it, seeing filmed sequences that replicate the actual views he would have were he in the space, driving. The experience of driving is made more intensive and involving through interaction: the user determines routes, turns, speeds, and points of view. He may also select the season, via a "season knob," and the visual mode of the tour: a photo, sketch, or animation (illustration 1). Thus the system does not simply repeat a guided tour, but allows a person to freely explore, at his own rate, via his own path, and with either photographic or detailed computer synthesized visuals.

A second purpose of the Movie-Map is to provide a data access and management system. The data stored may take the form of text, still pictures, synthesized or recorded sound, dynamic animation and cinema. It is characterized by being accessed spatially, where the particular organization corresponds to the physical layout of real space. In fact, this data is stored "behind" the facades of buildings, or in locales, and is retrieved by "driving" to it. Often the data relates exactly to the building which houses it: behind the front of a restaurant may lie the menu, or a tour of the kitchen. Other times, it may relate to the entire space and be located in an expected edifice: the telephone directory may be contained in the telephone exchange; the town's vital statistics are stored, naturally enough, at town hall.

A third function of the Movie-Map is to explore the interface between real images and computer graphics. The particular image recording and processing techniques used to create the map are uncommon, involving anamorphic lenses, travelling animation cameras, and three dimensional projection screens. Optical and electrical processing of the images, both prior to the recording on the disc, and during viewing are freely interchanged. With minimal real distortion and, ideally, no conceptual distortion, the user may place himself anywhere in the streets of the town, continuously alter his point and angle of view, and smoothly travel between actual camera positions. The ranges of realism necessary are being determined, and the means to achieve tham are being developed.

Two types of movie-maps will be discussed. The first, or the basic map, contains the elements necessary to implement surrogate travel and spatial data access. Travel is somewhat discontinuous, occurring in discrete movements along streets, and two views other than straight down the road are provided. This is the system on which many of the filming and mapping concepts were developed. It is characterized by being a single image display system, and no processing of the image occurs: appropriate images are selected from one optical videodisc at a time.

This will lead to the discussion of the anamorphic map, currently under development. This map is characterized by the addition of smooth travel between filmed images and continuous view alteration. The control systems for this map will prove to be very different, and the display system will use multiple screens, and curved screens. The images from the disc will be processed versions of the original film, and that processing will occur both before recording and during use.

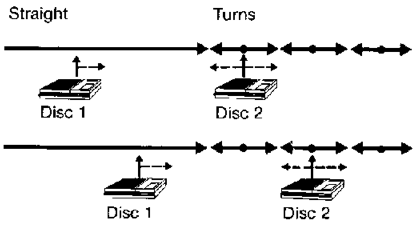
2.0 The Basic Mapping System

The basis of the movie-map is the optical videodisc. The capability of this device to store and randomly access 54,000 frames of television, and to sequence through them in a controlled manner is exploited heavily in the design of the system (see Appendix 1). Simply put, the map simulates travel by displaying to a user (driver) controlled rate sequences of indivisual television frames taken at periodic intervals along a particular street in a town. These frames originate as filmed views that correspond to what one would see were they actually in the town, driving along that street. Their rate of playback, or alternatively, the number of times each video frame is displayed before the next one is shown, determines the apparent speed of travel. To allow the route to deviate from straight paths down each street, separate sequences that depict all possible turns through all intersections are likewise filmed, and interspersed with the straight line travel as required. Thus, it is possible, for example, to drive three blocks along one street, at a set speed, then turn onto a crossing street and continue travelling.

Instantaneous freedom of choice in route selection is permitted by the use of two videodiscs. While one is playing the sequence that corresponds to travel along a street, the second is automatically positioned at the start of the sequences for turns from that street at each approaching intersection (figure 2). When a turn is requested, the signal from this second disc is made visible, and the turn is played. At the end of the turn, the first disc, having readied the sequence for the new street, continues the journey. This alternation between "visible" disc and "look-ahead" disc is the means by which all user choices in the system are executed. Note that due to the disc's ability to play backward as well as forward with equal facility, both a left and a right-hand turn may be queued on the look-ahead disc simultaneously. They are merely laid out "head-to-head," with one in reverse order: searching to their junction and playing forward displays one; playing backward displays the other (figure 3).

The configuration to realize basic surrogate travel and data access is shown in figure 4. A single, 32 bit minicomputer controls the operation of the map, accesses the data base, and manages the user interaction. Display is on a nineteen inch television screen, with a touch sensitive panel overlaid in front; images originate either in the videodiscs or in a frame buffer display, and are mixed, switched, and matted together by an interfaced video switcher. In the basic map, sound consists solely of synthesized speech, implemented by a Votrax, phonemic synthesizer. The user controls consist of the touch panel or a specially built "joystick."

During a drive through the town, a set of driving controls and indicators is overlaid onto the disc image, as shown in the figures. These pictograms define buttons and indicate the places on the screen which when touched cause a change in the action of the map. Centrally located is a stop button, which halts all travel. It is flanked by two colored bars that change either forward or reverse speed. When a speed is selected, a "tick mark" moves along the arrow to a point that is



Disc player alternation scheme. While disc 1 is playing a sequence corresponding to travel along a street, disc 2 searches for the turns approaching at the next intersection. (top). As disc 1 passes the intersection, disc 2 moves on to the next turns. (bottom).

Figure 2

Disc layout. A straight segment is followed by all of the turns from that street, in order. From a single point on the disc, playing forward will show a right turn; playing backward will show the corresponding left turn.

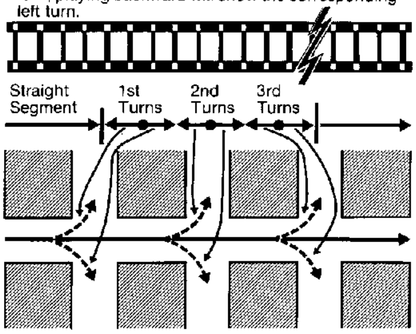
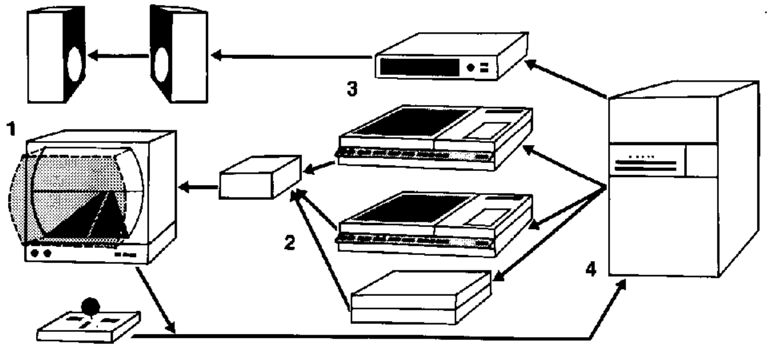


Figure 3



Basic Map Reading System: (1) User station with monitor; transparent touch panel; joystick; loudspeakers
(2) Video generation, two disc players; Ramtek graphic display, video switch
(3) Sound Generation, Votrax speech synthesizer
(4) System management, Interdata 7/32 processor

Figure 4

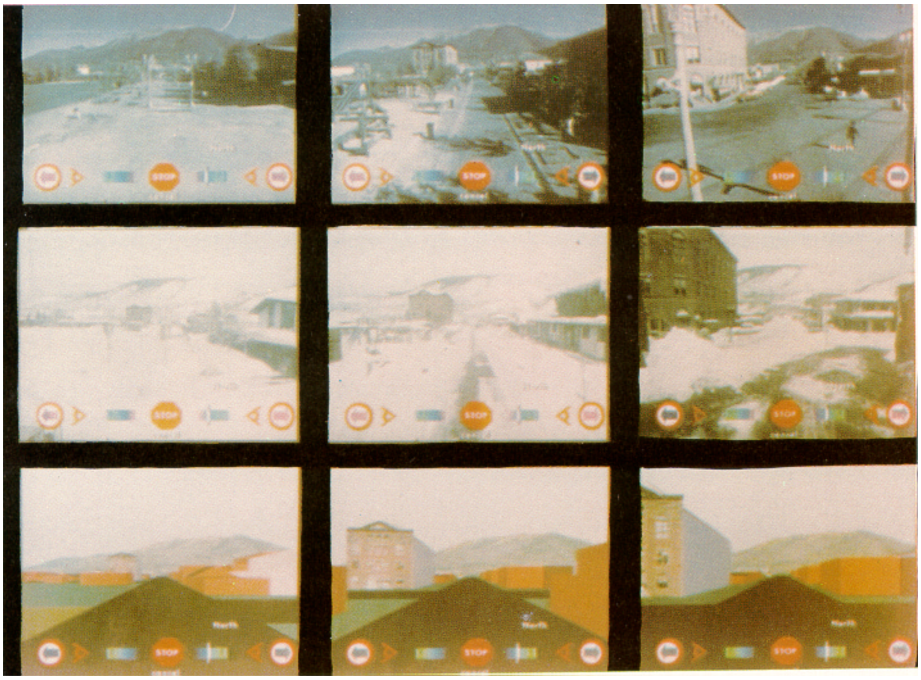


Figure 1
Surrogate travel shown in fall, winter and via animation.

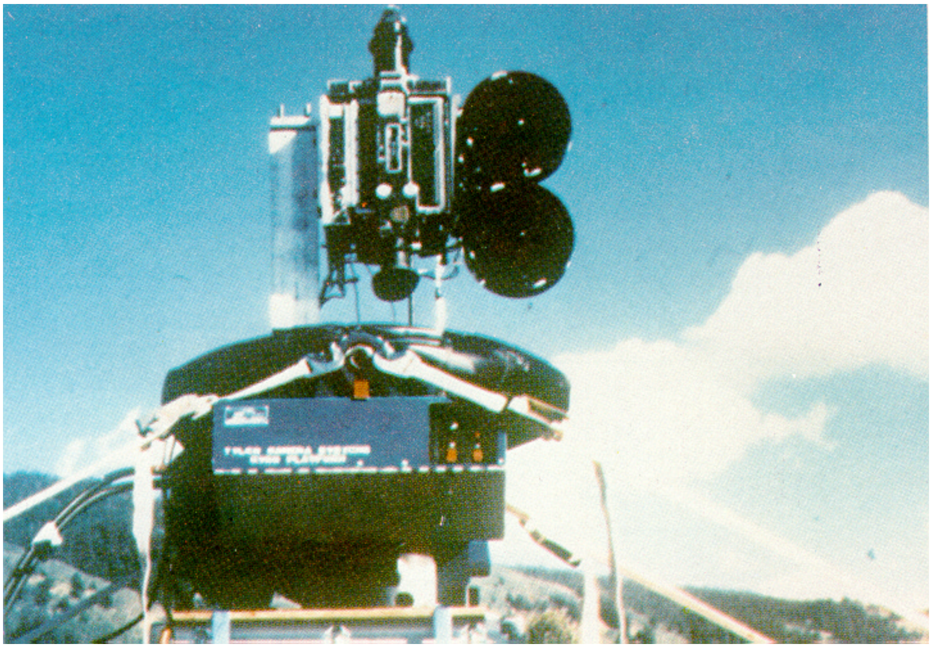


Figure 5
Stabilized camera rig, with 35mm single frame camera and anamorphic lens.

proportional to the actual speed chosen. When the user stops, that tick mark remains to facilitate resumption of motion with the same speed. At the extreme edges of the screen are two arrows which are abstractions of international road sign symbols for allowable turns. When touched, they change into green arrows to indicate that a turn at the next possible intersection will occur; when no turn is possible, a diagonal red bar is drawn through the turn indicator to replicate the "no left/right turn" international sign. Likewise, when forward travel is not possible, as in the case of a "T" intersection, the stop sign is replaced by an international "do not enter" sign.

Immediately to the inside of the turn indicators are two "eyeballs" by which a driver may request a change in direction of view. In the basic map, two alternative views exist: a full right or full left view; these are selected by the appropriate button. In the anamorphic map, the view rotates either clockwise or counter-clockwise.

All of the options available by touching the screen are similarly available by action on the joystick. Forward pressure controls forward speed, likewise for reverse, and sideways pressure indicates the desire to turn. Separate buttons radially spaced around the stick control are used for view selection. When driving with the joystick, there are two options available for speed control other than simple force-to-speed translation. The stick may be made to respond as an accelerator, in which case the rate of travel will increase in proportion to the time that the stick is pressed either forward or backward, as well as pressure, or the stick may be made to operate as a speed controller with coast. In this case, absolute speed is proportional to pressure, but when the stick is released, speed remains constant. When other than simple speed is the desired control, immediate stop can be effected by depression of the bottom button on the joystick, and gradual slowing can be accomplished by the reverse of the actions required to speed up.

In all cases, the indicators on the screen function to confirm that the system has received the command in question. Thus it is not necessary to hold the left turn indicator (for example) until a turn is begun, but only until the arrow responds. The next available turn will be made.

2.1 The visuals of driving

The town was mapped by mounting four cameras aimed at 90° intervals around a horizontal circle and taking one frame every ten feet along each street. Stability was maintained with a trailing "fifth wheel" that generated the shutter actuation pulses (figure 5). To minimize discontinuities such as lighting changes, abrupt shadow movement, and drastic alterations in the environment, filming was restricted to midday, and a path was developed that would completely record the images for each section of town in the least time. The entire town was recorded in both fall and winter; selection of image is done by a user "season" control.

The visual result of this filming technique is that in driving, one has the ability to stop every ten feet along the street, and proceed at rates that are independent of the rate at which the town was filmed. The actual speeds correspond to frame presentation rates that are multiples of 33 milliseconds, the normal video rate. When each frame is presented once, for 33 milliseconds, travel occurs at 300 feet per second, slightly above 200 miles per hour. In the actual system, frame rates that vary from ten per second down are allowed, and the speed range is from 68 miles per hour to a complete stop, with a granularity that is proportional to speed.

A corollary of the sparse sampling of images along the streets is that the image will often be viewed at faster rates than those at which it was taken (approximately two frames per second). This has the effect of translating orientation changes of the camera up in frequency, and thus amplifying their disquieting effect: a slow rotation will become a jarring bump. Also, since the motion is inherently abrupt, all other sources of abruptness must be eliminated if possible. This includes lighting and color changes, and shadow motions. If too many things vary between one frame and the next, a user can easily become lost travelling in a straight line!

As part of the work in developing the mapping technique, paths through the town were developed that approximated what filmmakers would call a "one take" film. The entire town was traversed in one pass, with the camera in use continuously from late morning until late afternoon. Edges of town were filmed either early or late, and north-south streets as closely as possible to their east-west cross streets.

Several different methods of filming turns through intersections were tested and incorporated into experimental versions of the map. They varied between "twirls" where the camera rotated through 360° horizontally in the center of each intersection, to complex paths that simulated the view a normal driver takes as he drives around a corner. In the one adopted, the camera was fixed to the filming truck and simply driven through all turns. While this requires eight passes through each intersection to record all possible street to street turns, efficient paths were developed to make this filming possible, and comparable in time to filming the straight sections.

2.2 The sound

During driving, all sound is generated from a speech synthesizer, and it is informational in content. The range of verbal accompaniment includes the following:

- Street one is on, approaching, or turning onto
- Compass directions
- Confirmation of turns, commands
- Spatial relationship to relevant, obvious landmarks
- Distance from known points

In addition, the verbalization of the above messages can be a function of the following:

- Real time since the last related message
- Real distance
- Number of previous repetitions of the message
- Intervening messages
- Intent in using the map

The program that controls the speech forms an ordered list of possible messages, then tests the entries in sequences and plays the first one that meets the weighted verbalization criteria. A user control varies the verbosity of the system.

2.3 Ancilliary data access

All data not related to actual driving is literally stored behind the facades of buildings and spaces. The images for these facades were recorded with specially calibrated single frame cameras that permitted accurate framing and seasonal registration between fall and winter. Thus the facades form a special set of pictures where the building to be examined is well centered in the frame and prominent, and where operation of the "season knob" results in a purely seasonal change; there is no movement in the image other than the disappearance of leaves and the appearance of snow (figure 6).

Since the facades are so evidently different from the normal travelling footage, they form a natural interface between driving and using the system as a means for accessing spatially stored data. To enter the data access mode, one touches the building in question during the drive. The normal driving sequence halts; the image of the facade is recalled from the disc. Simultaneously, the address or name of the building just touched is spoken by The synthesizer, and the driving controls are replaced by a set of indicators that inform the user of the name of the place and content of the available data. This new set of pictograms are touch sensitive and are used to select and control presentation of this data.

For most data access, the disc alternation scheme used in travel suffices. The facade image remains visible until the selected set of images is ready; then they are made visible. Two special cases are worthy of note. When the data requested consists of movies, the disc is used in normal play mode, and the movie can contain synchronously recorded sound. This is the only occasion in the basic mapping system where the discs play at normal rates and thus the only occasion where videodisc sound can be played without the use of a third player. When the data consists of textual information, that text may be generated from the stored data base and displayed via the graphics display. In that case, the discs are idle.

Data access mode is the least explored use of the movie-map, and as such, only a few examples of its operation are included in the system. Its controls are likewise primitive. The user may vary the rate and duration of a particular presentation in a linear fashion only, and stop it at any time to return either to the facade, or to the drive. All investigations of more sophisticated visual data control that might occur in the future are envisioned to use more than the two disc players that the basic map requires, and will allow continuous sound and more general perusal of ancilliary data.

2.4 Animation

Computer generated images provide the ability to drive in places where it is not possible to film, to generate an interactive overview of the town, and to remove excess detail that might confuse a newcomer. In one sense animation can be thought of as another season, and the selection of the animated footage is made by the appropriate operation of the "season knob." The animated tour is stored in the same manner as the filmed tour: with straight sections and turns, and operated likewise, with the standard driving controls. The difference is in the image content.

The approach taken to computer generated images is novel in two respects: it is created "off-line" and made "real-time" by use of the videodisc; it contains a mixture of two-dimensional and three-dimensional, synthesized images. The system used to display the pictures operates from a full 3-D data base of the town, in fact the same data base that provides the information necessary to allow one to touch buildings en route. It portrays each edifice as a simple rectangular solid, and its sorting is optimized by use of the block organization of the town. A depth sort is used, and the image is simply written to the display from far to near. In the course of writing the image, digitized facades are wrapped onto the surface of the solids, those facades

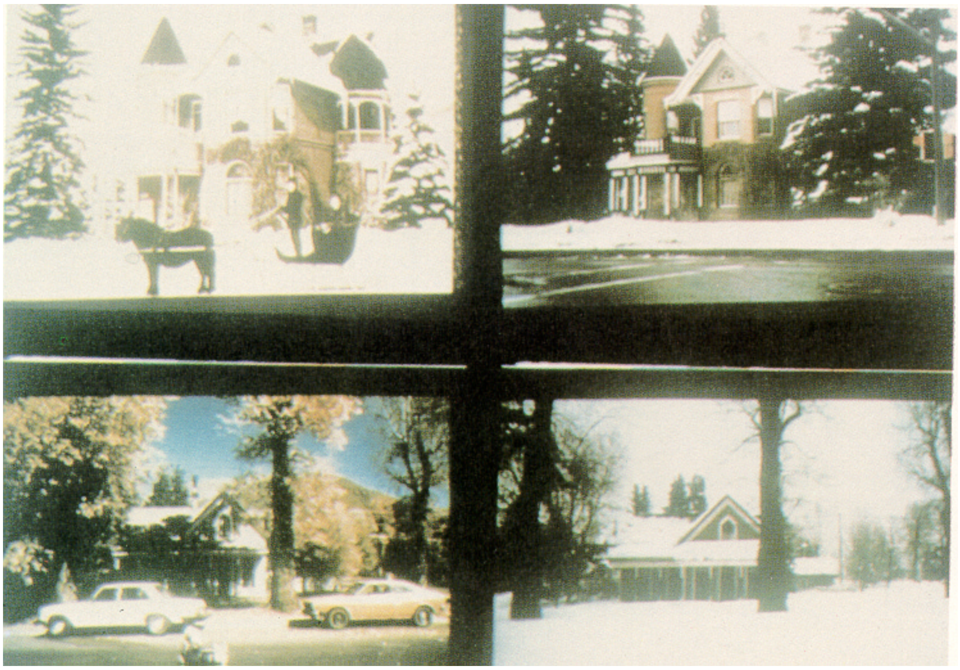


Figure 6

Historical juxtapositions (top)
Seasonal juxtapositions (bottom)

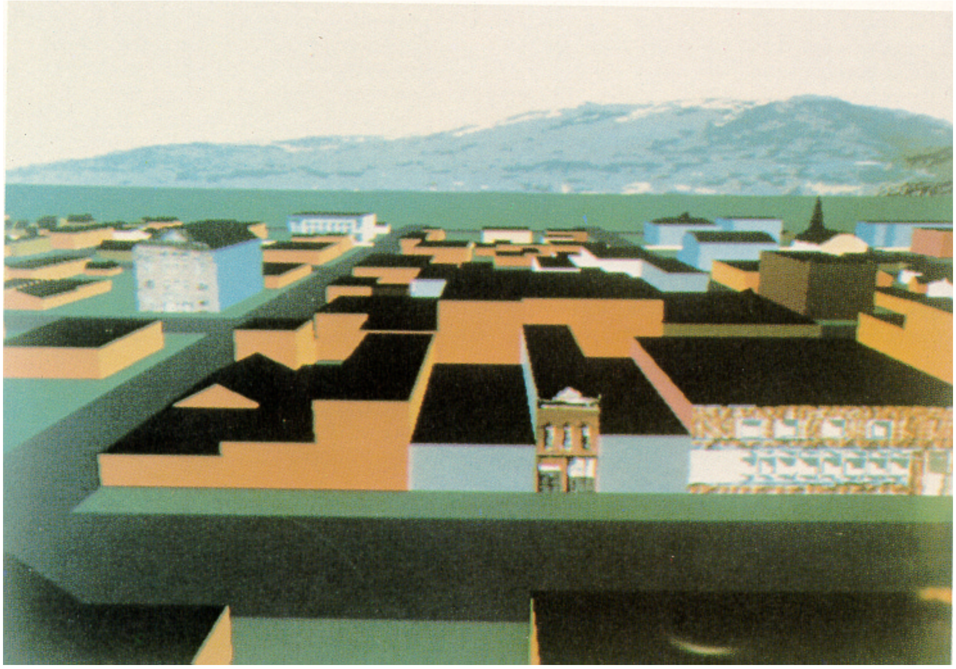


Figure 7

Animation showing "billboarded" facades and mountains.

being the same used in the real image sequences, and appear as "billboards." Distant mountains are treated similarly and are thus perspective renderings of two-dimensional images (figure 7).

In this manner, detail is added to the resultant image without incurring the overhead of a full three dimensional model of the building in question. Although the result is a scene that is not analytically correct, the effect is startling: even buildings with marquees that jut out into the street are readily identifiable and distinguishable. Further, the addition of mountains in the background serves as a valuable orientation aid and makes the image visually complete.

It should be noted that the animated version of the drive through the town is used in exactly the same manner as the real-image drive. Animation consists of forward travel segments and segregated turn sequences. They are concatenated into a continuous movie by the same disc-stepping and alternation scheme used with the photographed images. Thus animation that is created at less than real-time rates is used in an interactive, real-time manner.

Two types of animated drives are used in the mapping system. In one, the animation is literally a computer generated version of the real footage. Its point and angle of view duplicate it exactly. It is thus simply an alternative season and is used as such. In a second use, the point of view is elevated to make visible several adjacent streets. The drive is more like a helicopter tour.

Under investigation is the use of the animated images to provide intermediate vertical travel between a set of aerial photographs and the street. The procession might be as follows: one starts at the streets, travels upward to aerial height viewing computer generated images, then continues with photographic images taken from an aerial map. These allow the driver to ascend to heights from which it is possible to see the entire town and its surroundings, then return.

3.0 Anamorphic Mapping

The utility of the basic map rests upon two presumptions: that the impression of sufficiently smooth travel can be accomplished by the successive presentation of spatially sparse images, and that one can integrate those images to form a conceptual map of the town. Those assumptions can be thought of as goals, and the basic map is an attempt to make a minimal system to meet them. It uses only two videodiscs, and executes no image processing functions other than selection and sequencing. Investigation into anamorphic mapping extends these goals. The object is to make a system so immersive, so visually clear and simple, that the map will substitute for the first visit. A user will be able to form such a complete model of the town that he will arrive a native. This will be accomplished through enhanced responsiveness and compelling visuals that allow no doubts and misinterpretations.

The work being done falls into two general categories: optical processing and video rate electronic processing. All is being done within the context of available videodisc hardware; future work postulates modification of the disc itself to better integrate it into the processing environment. Initial experiments have been completed in the optical domain, and the first panoramic map exists along with a projection-based panoramic viewer. Additionally, processed images derived from the photographic footage and used to digitally synthesize turns is included in the system.

3.1 Optical processing

A tour through the town was filmed with a special panoramic filming lens that captured a 360° field of view with a single camera, and recorded that view on a single frame of 35mm film. The specific lens is related to a super-wide "fish-eye" type, with significant differences. A fish-eye lens accepts such a wide angle of view that it can "see" behind the camera. The standard of the industry is the Nikon 220° fish-eye. It maps an area larger than a hemisphere onto a circle on the film. When this lens is aimed upwards, a full 360° panorama is taken, from directly above to 20° below the horizon. However, most of the film area is occupied with the sky, and the buildings and streets all appear in a narrow ring around the outer edge. The actual lens used in the mapping eliminates the central sky area, compressing it into a smaller circle in the center of the frame, and maps approximately a 30° below horizontal and 30° above cylinder onto an annulus. Thus it is similar to a 240° fish-eye with the central 120° removed.

Images taken with this lens can be used in a variety of ways. The film can be "straightened" optically by projecting it back through the lens and re-photographing it with standard reproduction cameras, or linear panorama cameras. When used in this manner, the lens defers the decision about display camera angles to a later time or simply compensates for problems with moving lens panorama equipment.

A more interesting case obtains when the images are directly recorded from the film and viewed through a special viewer (see figure 8). With this viewing apparatus,

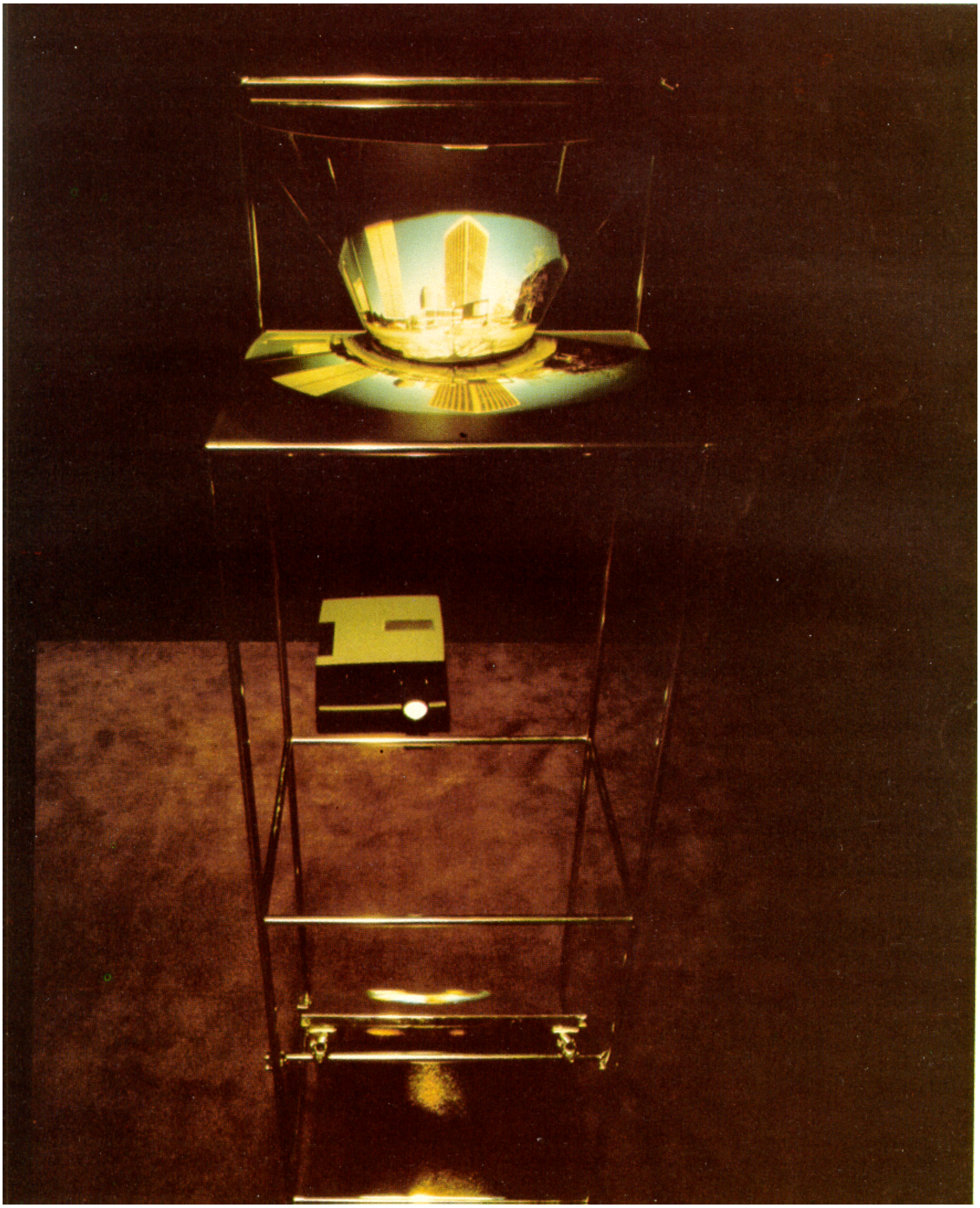


Figure 8

Optical System for viewing anamorphic images. Image is projected onto a horizontal screen with a conical mirror placed above. The image is viewed through the mirror.

each frame is similar to the anamorphic art works that enjoyed popularity in the renaissance. A distorted, incomprehensible view is rendered clear by the use of a conical mirror inverted above a horizontal display screen. Two benefits accrue. The user may "look to the side" by merely doing so; no digital controls are necessary, and the image is somewhat spatially faithful to the original scene. Further, the image is virtual, appearing in space and divorced from the screen, with all its fiducials and implied flatness; inherent spatiality is enhanced.

3.2 Electronic processing of the image

The addition of real-time video processing hardware between the disc and the viewing screen allows a plethora of enhancements to either the panoramically filmed map, or maps filmed with standard flat lenses. The range of complexity of this equipment is likewise variable.

The first step is to consider ways to present alternate views without disorienting the map reader. In the basic map, two 90° side views are instantly available, but to see them, an abrupt "cut" is made. In one frame the forward view is shown, in the next a sideways shot. This can foster some disorientation. A simple answer to this is to "slide" the side view across the screen, in the same manner as occurs during a video wipe. The effect of this will be a clear distortion in the imagery: the same building in both pictures will be simultaneously shown with grossly different perspectives. However, this joint presentation may will be conceptually clearer than the cut. The smoothness of the change will offset its inaccuracies. In this manner, a trade has been made between the physical error and the conceptual error.

The actual hardware needed to implement this type of transformation is relatively simple. A video delay with only a single line of storage will suffice. Two images are written into different parts of the lines, stored and read out as one. This is realizable digitally with little difficulty; the hardest part is the synchronization of the two videodiscs. Digital simulations of this technique are included on the disc.

An extension of this approach is to actually perform the perspective correction as the turn to the side view is done. To understand this, it is necessary first to understand that four cameras equipped with 90° horizontal field of view lenses correctly positioned can record all the data necessary to reconstruct that scene from any angle of view. To do so, four projectors are arranged in the same orientation as the cameras, and the images are projected onto screens that form the sides of a box. When a camera is positioned at the center, the perspective seen is an exact replication of what the camera would record were it in the original scene. It follows that a camera so positioned could re-record any angle of view, or an electronic processor can generate it from the original images.

It is therefore possible to extend the technique of wiping the side view into place with the technique of accurately rotating it into position. This requires an extension of the hardware to store more than one line of video information, and for it to be able to construct elements "in between" actual elements stored. It must now interpolate. Several examples of this "synthesized twirl" are used in the system (figure 9). They are created off-line and recorded with the animation camera.

One more extension to the use of digital hardware will now be made, capitalizing upon that interpolation capability. When the system can interpolate between picture elements and lines, it now can scale an image by an arbitrary factor. Thus it can perform the electronic equivalent to an "optical zoom." This technique can be used to smoothly interpolate between the successive frames taken at ten foot intervals to allow the user to position himself between actual filming points. A sequence of travel along a street can now be generated as if frames were shot at a spacing corresponding to that of travel.

This is another case of an actual perspective distortion that may result in a better conceptual presentation. The frames interpolated between the camera positions will not, strictly speaking, be accurate. This may be more than compensated for by the smoothness introduced into the act of traveling. This has not yet been attempted.

One final note on the use of digital hardware: the same processing equipment used to control rotations to side views may also be used to synthesize turns. A turn can be created that consists of a forward motion combined with a rotation of the camera. The camera continuoualy rotates from the straight ahead position until it points directly down the object street. Travel then proceeds in the new direction. This "side effect" of image processing greatly simplifies the on site work necessary to create a movie-map.

4.0 Conclusions

Although originally intended for the consumer market, the videodisc has proven to be a useful addition to the hardware of computer graphics and interactive systems. When regarded as a device to store 54,000 possibly unrelated single images rather than one half hour of continuous video,

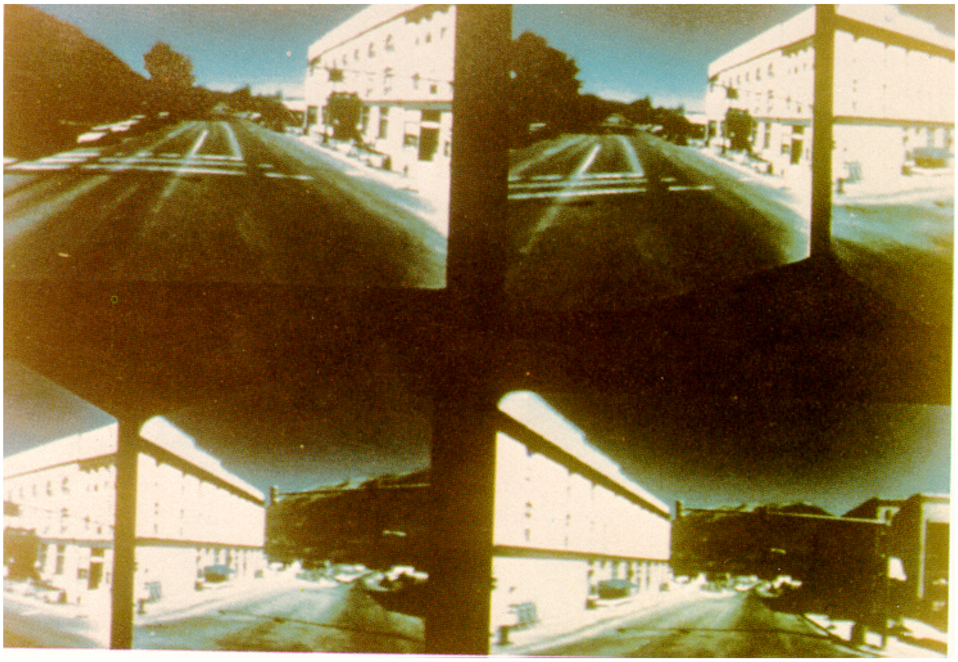


Figure 9

Computationally synthesized intermediate views, used to generate turns.

its power becomes obvious. The movie-map is merely an early attempt to use that power. Thus far, it has proven through informal testing to be a reasonable and useful replacement for a topographic map. Travelers have driven through the town in the lab, then visited the actual environment and found the map representation to be accurate and helpful.

The excursions into image processing also take on a different meaning in the context of the videodisc. When so many frames are available for processing, the importance of what can be done in real time must not be overlooked. The disc, when used in this type of environment can allow rapid processing and assembly of sequences, and changes the attitudes of graphics from one of making single images well, to making multiple images better and more efficiently. Future investigations will center on incorporating the dense spatially located video information into the image processing path. The disc will become an element in the chain, not merely the source.

Appendix 1: Videodisc operation

A videodisc is a record that stores television. An FM encoded signal that represents the video is stored as sequences of pits in a reflective plastic medium. On reading, the pits modulate a laser light beam, which is then decoded back to the original signal. In the recording scheme capitalized upon for the movie-map, there is one separately numbered frame per revolution of the disc; it may be repeated simply by the laser re-traversing the same "track." Slow motion and reverse are special cases of this repeat feature. Thus the disc can be used as a random access still picture store with the added ability to play consecutively stored pictures at periods that are multiples of the video frame rate with no interruption. In fact, the disc also can store two sound tracks along with the video, each of which may be separately enabled. Use of sound is possible, however, only in normal play mode, when the reading laser remains continuously in the "groove."

FOOTNOTES

1. Simulation of wrinkled surfaces, James Blinn, SIGGRAPH 1978.

2. An improved illumination model for shadow display, Turner Whitted, SIGGRAPH 1979.

3. Ibid.