EUROGRAPHICS '99 / P. Brunet and R. Scopigno
(Guest Editors)

Volume 18 (1999), Number 3

Experiments in Digital Television

Philipp Slusallek, Milton Chen, Brad Johanson

Computer Science Department, Stanford University

Abstract

Television has been around for many decades now and has become ubiquitous. Analog TV technology, however, which was developed before the 1940's, has remained essentially the same over the last 50 years. Alternative digital transmission standards have been developed recently and are in use for satellite and cable transmissions. This technology is now being deployed for terrestrial TV as well. This high-speed infrastructure for digital broadcasts of video, audio, and other data, which will be build over the next few years, suggests that we look more closely at this technology and explore its use from the standpoint of computer graphics.

This paper provides an overview of the technology of digital television, discusses its potential and challenges, and presents several experimental applications which demonstrate how computer graphics can provide input and influence the future of digital television. These applications include a virtual VCR, interactive video overlays, technology for the lecture of the future, and interactive video panoramas.

1. Introduction

Television (TV) has become a significant part of our life over the last few decades. Available in most households, it is not only a commodity item, which is — contrary to a computer — easily used by anyone, but also a primary source of entertainment and information.

One of the major characteristics of TV is its broadcast model of data distribution. In contrast to the Internet, TV is mainly used for providing content to large audiences. On account of this, production values are high and therefore content is expensive to create. Broadcast decisions are centralized and are only weakly influenced by the viewers through feedback on their channel selections. Interactivity in TV is strongly limited and requires a separate back channel — usually a telephone.

All this is in strong contrast to the Internet and the World Wide Web. The Internet provides peer-to-peer communication, where anyone can author his or her own content with interaction only limited by the available bandwidth. All this is attracting the interest of an increasingly large part of the public, although time spent using the Internet for entertainment and information retrieval is still tiny compared to the number of hours spent in front of TVs.

The Internet is not, however, well suited to the simultaneous distribution of large amounts of data to many places. Experiments with video-on-demand have mostly been failures, and the distribution of the Starr Report on the web almost brought the Internet to its knees. While we can expect new technologies, (such as multi-casting) and higher bandwidth, to alleviate some of these problems, the broadcast model of TV seems better suited to the delivery of large data sets to many different consumers (such as the morning news, stock tickers, etc.) than the Internet.

With the introduction of digital TV (DTV) [10, 16] a high bandwidth digital broadcast infrastructure (20-60 Mbit/s per DTV channel) is being set up in many countries around the world and will become available over the next few years. While it is designed for sending real-time video and audio, this digital infrastructure is equally well suited to other kinds of digital data transmission that are compatible with the broadcast model. In this context it is prudent to look into the possibilities and limitations the new medium offers to the computer graphics community and to explore where we can contribute to its development.

DTV and the broadcast infrastructure is an interesting new technology from a computer graphics point of view for several reasons:

- TV has always provided a particular successful form of "Virtual Reality". While not technically *immersive* like a CAVE [6] (in the sense of 3D immersion), TV allows the viewers to imagine themselves in another location by providing images and sounds that suggest the larger remote location. Certainly TV has been providing immersive experiences to more people than any other VR equipment to date, mainly by providing highly realistic images and convincing story telling. We believe that computer graphics can learn a lot from these concepts.

- The recent wave of image-based techniques promises to narrow the gap between purely image-based TV and computer graphics. Computer graphics offers many tools and techniques for extending the model of TV from real-time and isochronous video and audio transmission to other models for creating user experiences (such as textured 3D models, panoramas, animated sprites, etc.).
- Watching TV has always been a purely passive activity (aside from channel hopping). With the advent of DTV, new forms of interaction become possible. This is an area where computer graphics has a long tradition and can provide valuable contributions. These computer graphics techniques, however, need to be reviewed in the context of the mono-directional broadcast medium that is provided by DTV.
- TV sets and computers will technologically converge. DTV receivers require considerable computational power to receive, decode, and display the DTV program. It seems reasonable that some of these cycles can and will be used for other tasks such as displaying web pages or running interactive applications. On the other hand, computers are already able to act as DTV receivers (see Section 5.1). It remains to be seen exactly how this convergence will proceed, but both TV technology and computer graphics should be able provide significant contributions in any case.

In this paper we explore some of the options for DTV from a computer graphics perspective. The first part of this paper provides the basic background of DTV. We describe the basic technology and discuss current standards and systems. In the second part, we present four applications: virtual VCR, video overlays, lecture of the future, and video panoramas. These applications explore several aspects of the large design space that DTV offers such as surround video, interactivity, non-isochronous display, enhanced information display, and mixed-rate video streams. All applications have been developed as part of the ongoing *Stanford Immersive Television Project* and are running on experimental DTV-PCs. Our main intent with this paper is to spark interest for this interesting new area of research and to provide ideas and suggestions for further work.

2. Analog TV Technology

The ability to communicate with other people over long distances has been a desire throughout human history. Not long after Morse invented the electric telegraph, people suggested that images could also be transmitted electrically. The development of television was a long struggle starting with

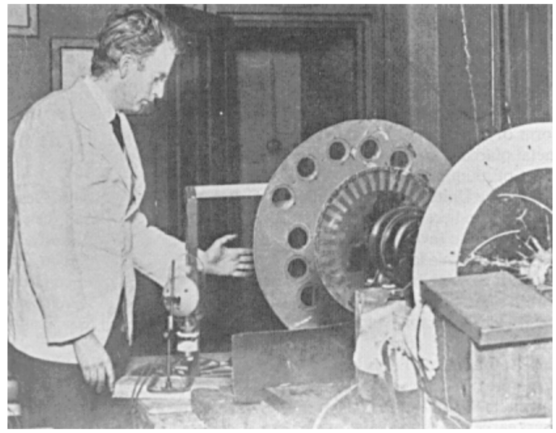

Figure 1: *John L. Baird, one of the TV pioneers, and his mechanical TV in 1925. It provides only a few scan lines and could only transmit crude outlines, but it already used the basic principles of analog TV transmission.*

the first experiments around 1880. However, it was not until early this century that inventors overcame the immense technical problems (see Figure 1). It required a tremendous scientific, technological, and industrial effort before the first, reliably working electronic television cameras and receivers were built in the 1930s and the first TV standard (NTSC) was defined in 1941 [1, 9].

Essentially, a television camera scans an image using horizontal lines, converting the light intensities along the way into electrical signals. In black and white TV this signal is directly modulated on an RF carrier after it has been augmented with synchronization information. On the receiver side the signal is demodulated, and displayed by synchronously scanning the screen phosphor with an electron beam replicating the observed intensity pattern.

Even for black and white TV, compression by interlaced scanning had to be used to squeeze the immense data stream into a narrow RF band. Interlaced scanning allows the refresh rate to be doubled to the ergonomic 60 Hz by sending only every other line per screen update. The addition of color to TV in the 1950's introduced even more bandwidth problems. They were solved by significantly compressing the color signal without too much loss in color fidelity by exploiting limitations in human visual perception. The combined black and white and color signals still had to fit into the same small RF bandwidth of 6-8 MHz, leaving little room for any further extensions or improvements to the standard.

As a result, the technology of the initial TV standard, while brilliant for its time, severely limited the quality of TV signals for many years to come. The NTSC standard in particular suffered from this technology limitation. Since the 1940's the basic technology of TV has remained essentially the same, although the equipments themselves have improved dramatically. Given the newness of the technology and the dramatic competition and pressure under which early TV was standardized and marketed, it is amazing that we have been able to get by with the same TV technology ever since.

Nonetheless, the (lack of) quality of video and audio has been a constant nuisance and many improvements have been suggested (such as the many forms of High-Definition TV (HDTV)). The main problem for all of them – and one of the major reasons why little has changed for TV for so long – is the analog nature of TV. Any improvement to the image and sound quality would have required higher bandwidth, which was very difficult to arrange given the huge existing TV infrastructure.

Analog HDTV systems that would offer significantly better image and audio quality have been proposed in many variants over the last two decades. They always suffered from the bandwidth problem, because of the difficulty in efficiently compressing the analog signal†. It took considerable time for the (often strictly analog) TV community to realize that only digital processing could provide the necessary compression ratios to make HDTV practical.

3. Digital TV Technology

A major limitation of analog TV processing is the conversion of the video image into a serial stream by horizontal scanning. This makes it hard to perform any vertical processing on the signal. In order to compress a video image well, both horizontal as well as vertical processing is required. Video compression also relies on taking temporal coherence into account, by comparing successive frames and only sending the difference information. All of those computations are greatly simplified by breaking the picture into blocks of 8×8 or 16×16 pixels each, which can easily be done in a digital representation.

For digital broadcasts, the video and audio streams are separately compressed, split into packages of a fixed length, and then multiplexed into a single digital MPEG transport stream [10, 13, 16]. Before the DTV transport stream is broadcasted, it is augmented with forward error correction and is modulated on an RF carrier. Depending on the DTV and transmission system, each DTV channel offers a payload bandwidth of roughly 20 Mbit/s for terrestrial, 30 Mbit/s on cable, and 40-60 Mbit/s using satellite broadcasting. This is enough for distributing at least one HDTV or 5-8 standard TV programs.

In addition to audio and video data, the MPEG transport stream can also contain other data streams. Most important is the *program information stream* that tells a DTV decoder which video and audio streams form one program and should be decoded and rendered together. Auxiliary streams such as multi-language audio, closed caption, teletext, and other arbitrary data can also be included in a DTV stream.

The main benefits of a DTV system are:

Image and Sound Quality Due to digital error protection and recovery, images contain no snow or other analog noise. However, in cases of significant compression or packet loss, blocking artifacts and ringing can occur (well known from JPEG-compressed images). Similar effects apply to sound.

Higher Resolution DTV allows the transmission of higher resolution images (up to 1920×1080 interlaced). It turns out that the increased image quality at standard resolution (480 lines of video for NTSC) in progressive scan mode (i.e. without the interlacing artifacts) is already considered a significant quality improvement by most viewers [19]. It remains unclear for now which resolution is sufficient for being considered HDTV. One drawback of HDTV programs is that it requires the consumer to buy expensive high-resolution displays to take full advantage of the higher resolution.

Extendibility Due to the very general MPEG transport stream, any new services can be easily added to a DTV system without compromising existing DTV receivers. Also, better video and audio compression technology installed on the head ends promises to transparently make more bandwidth available to other services in the future. New developments in scalable video coding even make it possible to incrementally add the missing pixels of HDTV to existing low resolution video streams [7, 10].

Data Casting A digital TV system has the other important advantage in that essentially any data can be broadcasted (data casting) — video and audio being only the two most obvious examples. Examples of data casting are simple broadcasting of data files, auxiliary data to a TV show (like overlays and event-driven scripts controlling their appearance, see Section 5.3), and collections of video and audio streams that together make up a new TV experience (e.g. video panoramas, see Section 5.5).
One very simple extension of DTV would be to use any excess bandwidth (e.g. at off prime time hours), or even whole DTV channels to broadcast data such as web pages, news, software updates, etc. DTV receivers with disks could store or cache some of the data depending on user preference or subscription and make it available on demand. It seems that data casting will become an important aspect of DTV networks, although it is yet unclear which of these services will be commercially successful.

Although proprietary DTV systems have been in use for some time in specific satellite systems (e.g. DigiCipher II), the two major DTV systems now are the European DVB (Digital Video Broadcast) [8] and the American ATSC (Advanced Television Systems Committee) systems [2]. DTV was first introduced for satellite transmission and is only now

† The Japanese HiVision/MUSE standard actually uses a number of DSPs to perform signal processing on the analog TV signal.

also being deployed for terrestrial broadcasting where it is supposed to replace analog TV over the next decade or so.

3.1. The DVB System

DVB has been developed as a consistent base standard for the DTV stream that is augmented with standards that specify channel coding and forward error correction suitable to the specific characteristics of transmission channels such as satellite (DVB-S), cable (DVB-C), and terrestrial broadcasting (DVB-T). In particular, the modulation scheme (COFDM) for terrestrial broadcasting supports mobile reception and is well suited to difficult signal conditions (e.g. ghosting) requiring only small settop antennas. Early on the DVB consortium included facilities for data broadcasting, interactive services, and conditional access into their standards.

For the DVB consortium HDTV had low priority on the list of features and has only recently been added to the standard [17]. The DVB consortium plans to simulcast separate standard and high-definition video streams instead of requiring each receiver to be able to decode and down-convert HDTV streams. Because conditional access is well-defined for DVB, the same receiver boxes can be used for free channels as well as for pay-per-view services by inserting a specific decryption PCMCIA/PC-Card into the box. Many subsidized DTV receivers in the UK are now distributed at low-cost ($300) by pay-per-view companies making DTV affordable to the consumer.

DVB already is the major standard for most satellite systems and many cable TV systems. Terrestrial broadcasting using DVB started in the UK in last November, with Sweden and Australia soon to follow. Other countries still rely on satellite distribution of DTV, but most plan to introduce terrestrial DTV in the near future.

3.2. The ATSC System

The American ATSC system started off by concentrating on terrestrial broadcasting only. Because HDTV has been a major motivation for the introduction of DTV in the US, its distribution is actually mandated by the US government. However, stations plan on broadcasting at most 3% of the daily programming in HDTV.

Interestingly, the FCC adopted ATSC standard does not define a specific video format or resolution, leaving this choice to the TV networks and receiver manufacturers. The consumer, however, is left with the possibility that his new receiver may not actually be able to decode all available programming. It still remains unclear how well DTV receivers will be accepted by consumers, since their prices in the US are likely to be high, mainly since each box needs to be able to decode and possibly down-sample HDTV programming. It appears that some manufactures will ignore HDTV in favor of low cost receivers.

Although both standards are based on similar concepts, there are enough differences that make DVB and ATSC incompatible:

Modulation Each system uses its own modulation and error correction scheme for terrestrial broadcasting.

Program Information The data streams that describe the available programs are completely incompatible, so that neither system would find the other's programs even though both understand the same MPEG transport stream and video compression syntax.

Audio Different audio encoding schemes are being used. DVB uses MPEG Layer 3, while ATSC uses the Dolby AC3 standard, although both standards allow the same 5.1 channel, CD quality surround sound [10, 12].

Video Although the MPEG-2 compression scheme [10] is the same, both standards have chosen different MPEG parameters. This may render the video streams incompatible with decoders that are specifically designed for one system or the other.

Due to market constraints it currently seems unlikely that the two standards will eventually converge, although manufacturers might build dual standard receivers in the future. Unfortunately, Japan seems to be determined to develop yet another DTV standard – ISTV (Integrated System TV) – in addition to its current analog HiVision/MUSE system. This may introduce yet another standard into the mix.

Many TV stations and networks see DTV as an opportunity to gain new market share. The situation is particularly problematic in the US, because the introduction and continued support of an HDTV capable ATSC system has cost US TV stations billions of dollars for new equipment and distribution infrastructure. The stations are now looking for new sources of revenues to make up for this loss. Although for each analog channel a TV station received an additional free DTV channel if it promised to transmit HDTV programming, it seems highly unlikely that a larger number channels can cover the costs of DTV. As a result, many TV networks and stations are looking into adding simple data casting to DTV and augmenting their DTV program with new user experiences through interactive components and enhanced visual presentations.

4. DTV and Computer Graphics

In the following section we explore some of the interesting dimensions of the DTV design space, such as the time dimension where the presentation time can be decoupled from the transmission time, the presentation of additional on-screen information linked to the DTV program, new user experiences through extended and semi-immersive visual presentation, and finally the interactive dimension of the previously passive medium of TV. In particular we concentrate on areas where computer graphics offers techniques and experience for future DTV applications. Concrete application scenarios and example implementations are then described in Section 5.

4.1. Time Warping

A DTV receiver with enough disk space could easily act as an intelligent VCR or a video buffer, that can – automatically or on command – save your favorite shows and programs to disk if you happen not to be at home. Current 18 GB disks drives can provide up to 10 hours of video depending on compression.

The strict isochronous nature of watching TV can be eliminated by temporarily buffering the incoming video stream of a program on a hard disk, while you pick up the phone or prepare a snack. The paused program can then be continued from the stopped location running off the hard disk, while the incoming video continues to be buffered until you have a chance to catch up by skipping the next commercial (see also Section 5.2). This new technique requires little support from a computer graphics perspective except for very simple and intuitive user interfaces, but offers a basic service for many of the higher level applications described below.

4.2. Video Overlays and Insets

There are many occasions where a viewer might want to get more information than is normally available in the video stream. A particularly good example is sports programming, where some people would like to see more detailed information about a game or a player (biography, statistics, related events, merchandise, etc.). This information can be provided in video overlays, which can pop up automatically or on the request of a user. The user can then choose to browse the information further, adjust its display, or dismiss it. The overlays can be animated or static, linked to hot spots on the screen (e.g. to where a particular player is currently visible), or could be textured insets attached to objects visible in the scene (e.g. advertisements on the floor of a basketball court). These latter forms of display require non-trivial image-based techniques both from the field of computer vision and computer graphics to properly operate and synchronize with the program.

4.3. Extended Video Presentation

In a traditional TV program every detail of the presentation is predefined by the TV station or network. This is a direct result of the chosen representation as a fixed, full-screen video stream. As soon as we extend the representation of video, we offer more options for the user to customize the presentation according to his requirements and preferences.

One general option that is currently being standardized as part of MPEG-4 is to split the video program into multiple “Video Objects” (VOB), which need not have rectangular shape and may be partly transparent. The VOBs are transmitted separately across the same DTV stream together with a default presentation arrangement, which may be three-dimensional. Depending on the parameters, the user may then be able to interact and rearrange this default presentation or substitute his own. Using these principles we have implemented a video panorama, which is described in Section 5.5.

Another opportunity that is offered by this framework is the adaptive coding of the different video streams, which may have completely different coding and update requirements. As an example, when transmitting a lecture the image of a blackboard requires very high resolution to be readable, but only certain regions need to be update at frequent intervals (see Section 5.4).

4.4. Interactivity

The video presentation outlined in the previous section already allows the user a limited form of interaction with the DTV program, but more degrees of freedom are certainly desirable. A major stumbling block to interactivity is the strict broadcast model of DTV, which does not offer an integrated back channel. Adding Internet access solves this issue and offers all the usual interaction capabilities. However, it is interesting to explore the available options without adding a back channel.

In this mode, only local interactivity can be provided. This interaction is limited to an environment predefined by the creator of the program. Examples are downloaded scripts that react to certain interactions by displaying, hiding, or modifying previously downloaded media presentations and program state. An example application is given in Section 5.3, where overlays can react to user events as well as time events coming in from the transmission channel.

4.5. Image-based Modeling and Rendering

While traditional geometry-based approaches can be applied to DTV, image-based algorithms seem to be particularly well suited to the task. We can often directly use input video streams and process them appropriately to produce new and enhanced DTV experiences as in the panorama example (see Section 5.5). Sending a “video” stream of range images or downloading a layered-depth images [18] would allow a client to perform local interactions with the DTV content.

4.6. Execution Environment

An important aspect for developing new DTV applications is the environment that is provided by a DTV receiver for executing downloaded software. The vendor neutral DAVIC consortium is working on API interfaces for interactive TV based mainly on the DVB standards, but other competing solutions exist as well.

It still remains unclear what support the execution environment will provide to potential DTV applications. How much CPU power and graphics support are required and how much can we possibly expect to see in a DTV receiver in the future? How closely will a TV set be integrated with the computer or will they eventually be the same server device with multiple display across the house?

It seems that as we come up with new services, the TV set of the future will need to adjust and expand accordingly and may need to be periodically upgraded as is necessary with computers today.

5. Experiments in DTV

In the following sections we describe ongoing work on a number of experimental DTV applications that explore the design space of DTV. In choosing what to persue we focused more on developing new experiences for the potential user than on direct applicability in todays DTV environment. Nonetheless, all the applications have been implemented within an experimental DTV infrastructure and have been tested in a broadcast setup.

5.1. Infrastructure

The infrastructure for our experiments consists of several 350 MHz Pentium-II PCs running Windows 95 with video boards that provide color space conversion. In addition each system has an ATSC receiver card (see Figure 2), which has been designed by INTEL's MicroComputer Research Lab. This card provides the basic functionality for receiving an ATSC modulated RF signal off the air via a standard rooftop antenna and perform channel tuning, demodulation, and error correction. The DTV hardware delivers the raw 19.2 Mbit/s MPEG-2 transport stream into system memory. Any further processing is performed in software.

The DTV hardware is accessed through an experimental API that shields the application software from the hardware details. At the lowest level this API demultiplexes the transport stream and delivers packets of chosen audio, video, or data streams. Layered on top of this interface are services for decoding and/or displaying a video stream, as well as receiving data objects (essentially files with identification information).

Since the hardware only provides the raw MPEG-2 transport stream, all demultiplexing and MPEG decoding is performed in software. Our systems allow us to receive and display a standard TV video stream (720×480, interlaced) in real time. Decoding of HDTV streams is an area of ongoing development.

An experimental ATSC transmitted set up by INTEL some 10 miles from Stanford has been used to provide the DTV streams. During development the DTV receiver card can simulate the reception of a DTV stream by reading the transport stream from a local file.

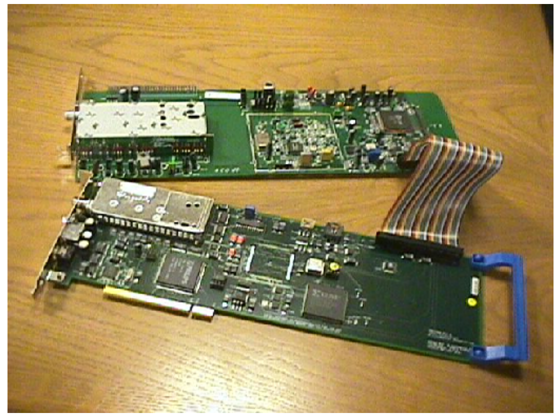

Figure 2: *Photo of digital/analog TV receiver cards build by Intel's MicroComputer Research Lab. The top board can receive NTSC signals while the bottom board handles digital ATSC signals.*

5.2. Virtual VCR

As part of the support infrastructure we have implemented a video buffer that supports VCR-like operations such as slow motion, pause, replay, and fast forward. The buffer could be used in applications such as personal sports replays, self-paced lectures (see Section 5.4), flexible guided tours (see Section 5.5), or for downloading a movie during the night that you want to watch tomorrow.

To implement the VCR-like interactive features, a cache manager is required that is able to buffer a large amount of data and supply it to the decoder in real-time. Since the memory requirement can be huge, a main-memory-only buffering approach may be prohibitively costly. The alternative is using a memory-disk integrated cache (MEDIC) [3]. Since the per-byte disk cost is about one hundredth of the per-byte memory cost, MEDIC is economically attractive.

MEDIC carefully allocates a limited amount of memory to competing tasks, i.e. to receiving new data from network channels, to writing data to disk as memory fills up, to reading data from disk as needed, and to holding data for decoding and playback. Since data is concurrently written to the disk cache and read from the disk cache, MEDIC must intelligently issue IOs to maximize throughput and to avoid undue conflicts.

With less than 8 MB of RAM our video buffer can support the interactive operations discussed above without causing any "jitter". This result confirms the theoretical study [3]. Figure 3 and the accompanying video shows an application of video buffering for implementing a Virtual VCR.

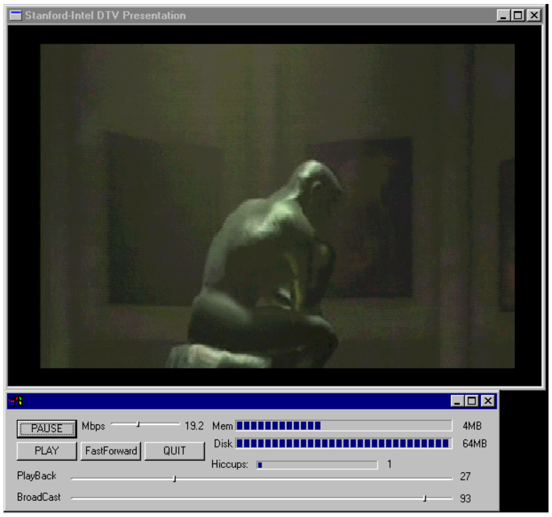

Figure 3: *A screen shot of the Virtual VCR, where the top window shows the decoded video. The bottom window shows the control panel, the current memory and disk usage, and the reception and decoding progress.*

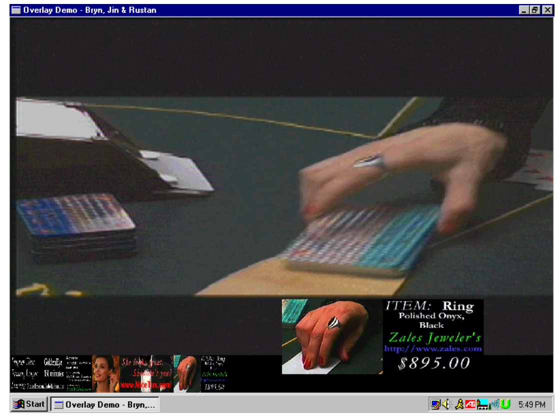

Figure 4: *A screen shot of the video overlay project. The feature movie is displayed in the center of the screen. A list of the available overlays (movie description, main actress biography, tuxedo advertising, and ring advertising) are shown on the bottom of the screen. The ring advertising overlay has just arrived (displayed in lower right corner). This context sensitive advertising is triggered by the movie scene that shows a similar ring.*

5.3. Video Overlays

The value of video overlays is readily demonstrated by the stock tickers, channel logos, and sports score boards that are already found in today's analog TV programs. In contrast to analog TV, DTV allows these overlays to be composited in the viewers' homes rather than in the production studios, thus allowing flexibility in the storage and presentation of the contents as well as customization of the viewing experience. The goal of the overlay project is to create an infrastructure for experimenting with overlays as well as example applications.

Our overlays are described by a markup language similar to HTML. In addition to standard fields, each object descriptor contains a time field. This allows the specification of not only where and how objects should appear but also when. The degree to which viewers can change the visual properties of the overlays can also be specified. For example, an overlay's position and size can be either fixed or customizable by the viewer. Each overlay object as well as the markup language is sent through a separate DTV stream. The overlay receiver handles four tasks: receiving objects, parsing markup language, reacting to viewer input, and displaying the overlays. When a customizable overlay is received, an icon is displayed in a sliding window at the bottom of the display. The user can then select this overlay to show or hide it, as well as position and resize it.

An important new use of overlays is context based advertising. In our video, we show two examples where the information for products in a feature film is made available to the viewers at an opportune moment. Such advertising, coupled with the Virtual VCR technology, will allow the viewers to pause the movie, consider and possibly purchase the product, and then enjoy the rest of the movie. Such advertising is potentially less distracting, more enjoyable, and more appropriate that current advertisements that completely interrupt the program. See Figure 4 and the accompanying video for an example of video overlays for DTV.

5.4. Lecture of the Future

Stanford University has a long tradition of transmitting many of its lectures on a specialized network known as the Stanford Instructional Television Network to which many of the nearby Silicon Valley companies subscribe. Most transmitted lectures consist of videos that are switched between showing the lecturer, the blackboard, and slides or other material along with audio. There are several problems with this style of presentation:

- The viewer can only see what the camera operator chooses to transmit. It is impossible to look at some material in more detail.
- The resolution of the blackboard image is often less than adequate, rendering some text unreadable. A higher resolution blackboard image would be of great help. Since its contents do not change very often, even high-resolution images would require little bandwidth.
- For long periods of the lecture, programs only show the talking lecturer. A lot of bandwidth is wasted since the background is usually not changing.

In our project we addressed these problems from multiple directions applying the enhanced capabilities of DTV and image-based techniques. First of all, we have chosen to separately process and transmit the image of the lecturer, the room, the blackboard, and any additional material. Each of these source materials have very distinct video characteristics that we plan to utilize. In the following we give a short overview of the whole project, but concentrating on the extraction of a high-resolution blackboard image.

On the head-end we need to capture multiple video streams of the room with the lecture and the blackboard. We then need to segment and broadcast the different objects (much in the spirit of MPEG-4's video objects [14]). On the receiver side we need to recompose the different video streams into a single presentation, but are now able to give the user the ability to customize it according to his preferences. This allows him, for instance, to concentrate longer on a blackboard image or review some earlier slides again.

We start by creating a geometric model of the lecture hall, which is augmented with projective textures extracted occasionally from a video stream. As a result we can save considerable bandwidth by only transmitting this model and the textures infrequently instead of sending the image of the background for each video frame. This approach allows a viewer to freely move within the room and view the classroom from whatever location he prefers, not just the angle chosen by the camera operator.

In order to display the lecturer within this model, we need to segment him from the background. We currently use a simple segmentation algorithm that uses the known colors of the background to distinguish it from the lecturer. Using the known camera position and the geometry of the room, we roughly estimate the position of the lecturer in front of the blackboard and place his video image as a 3D billboard into the scene. Although this is a simple technique it already provides a surprisingly realistic view of the lecture while using only a fraction of the bandwidth a full video transmission of the lecture would require.

In order to display the blackboard with high enough resolution to be readable, it is necessary to use several cameras to form a running image of the blackboard that is updated in real time. Since a single camera cannot capture the entire board with sufficient resolution, we use cameras that pan and zoom to areas of interest and integrate that data into the running high-resolution image of the board. A single fixed camera is used to obtain a low resolution reference image of the entire board to aid in the integration of the image streams from the higher resolution cameras. It also insures that something can be said about all of the board in the case that an area has not yet been scanned by one of the high-resolution cameras.

The lecturer is segmented from the low resolution camera's input to obtain the running reference image of the board, without the lecturer obscuring the view. The segmentation problem here is simpler than the previous one, since we are eliminating the lecturer, rather than extracting him and a conservative algorithm can be used that might also remove a small border around him. The remainder of the frame is then copied over the running low resolution image. To detect the lecturer, we threshold the intensity difference of the running image and the next video frame based on the fact that the blackboard will stay nearly the same. This simple technique can fail and is therefore augmented with a more robust but slower algorithm that analyzes the color distribution for those areas that have not been updated for a while (because we might have wrongly identified a piece of blackboard as a lecturer).

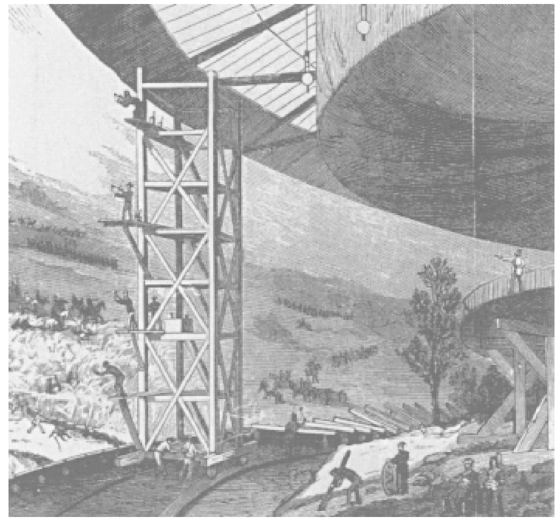

Figure 5: *Interior view of a panorama shortly before completion of the painting* [15].

To decide where to point the pan and zoom cameras, we maintain a "curiosity" bitmap. It is marked when we see a large enough difference in the corresponding pixel in the low resolution control image (which is updated in real time, regardless of the positions of the mobile cameras). The moving camera will then sweep out that area of the image, take high-resolution images, and clear the appropriate areas in the curiosity bitmap.

When integrating a high-resolution image from a high-resolution camera into the output stream, we compute the mapping between the camera's image space and board space by identifying markers on the board. After reprojection of the high-resolution image into the blackboard image it is masked with the lecturer and copied into the output image. Having a reference stream with the entire board is crucial to integrating the high-resolution images taken from arbitrary positions.

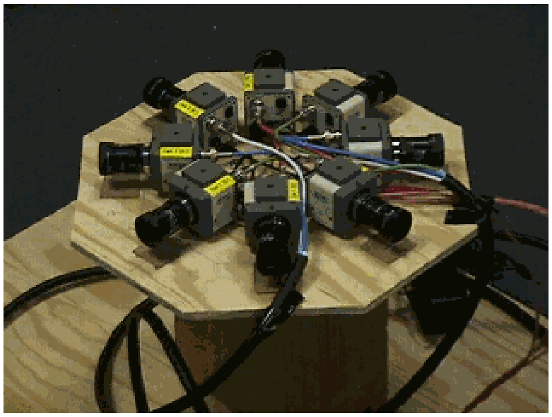

Figure 6: *The camera rig used for capturing the video panoramas.*

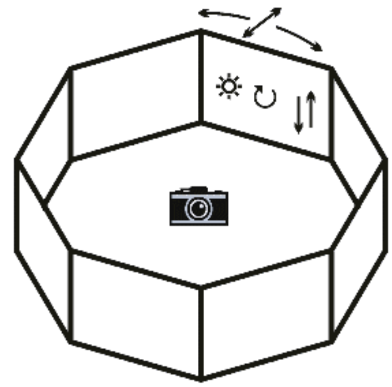

Figure 7: *The arrangements of the polygons on to which the eight DTV streams get projected. By adjusting the position, orientation, and brightness of the polygons, we calibrate the panorama to show a seamless surround view.*

We currently use a modified MPEG encoder to compress the high resolution blackboard image using a significantly lower frame rate. Ideally, we would like to only transmit those areas that have changed, but MPEG already has a fairly small overhead for coding these unchanged regions.

5.5. Video-Panorama

In the late 18th and 19th century, panoramas were a popular form of mass entertainment. Panoramas were invented by Robert Barker in 1792, as a means to show complete 360 degree views of interesting environments (such as landscapes) to the interested and paying public [15]. Panoramas in these days were created in specially build circular buildings with a small viewing platform in the middle and a painting on a large cylindrical canvas around it. These panorama canvases reached heights of up to 18m and measured up to 130m in circumference. In order to achieve the correct perspective view, the painting (which often took months to complete) had to be warped accordingly (see Figure 5).

Due to the large highly realistic drawings at a large distance the viewer, limited vertical viewing, and limited movements on the viewing platform, the human senses were tricked into believing they were seeing the real landscape. This experience was often enhanced by placing real 3D objects in front of the image that merged into the background (providing otherwise non-existent parallax information) and by adjusting the lighting conditions accordingly. In that sense, panoramas provided a early form of *virtual reality* experience and caused dramatic responses by the viewers of its time.

Animated “video panoramas” were first created by Raoul Brimoin-Sanson (Cineoramas) in 1897 and are still used today, e.g. in Disneys CircleVision installations. 3D IMAX and OMNIMAX [11] are limited forms of panoramas but also include stereo effects.

Panoramas recently got a new push in the context of image-based rendering by the work of Chen and Williams [5, 4]. Here, a cylindrical panorama is created and a small view can be displayed and moved interactively on a monitor, providing some sense of the surround view of a real panorama. These panoramas, however, are limited to static views points.

Using the infrastructure of DTV, we decided to tackle the problem of video panoramas by providing the viewer with an interactive view of a moving panorama. The context of this application would be a guided tour of the Stanford campus. At each instant the video panorama would provide a 360 degree surround view from the current location. As the location or surroundings changes, new panoramas become visible. The user should also be able to interactively move the current view within the full panorama, so he can be watching a particular object longer than the guide anticipated or he can simply enjoy the surrounding campus.

In order to create the 360 degree panorama we had to capture live video from eight CCD cameras arranged on a circular platform. In order to manage the immense video bandwidth of roughly 240 Mbit, two quad video mergers are used that combined the analog signals from the eight cameras into two NTSC video signals, each divided into quadrants showing one of the cameras. These two video signals are captured at 752x480 resolution using Motion-JPEG capture boards on two separate PCs. Finally, the eight video streams are extracted from the captured footage, encoded using MPEG-2, and multiplexed into a single ATSC DTV channel.

On the receiver side, a pipeline of programs work together to display this transport stream. The DTV stream is received on a PC, where it is retransmitted over 100 Mbps Ethernet to an SGI Onyx 2 with eight CPUs. The standard reference

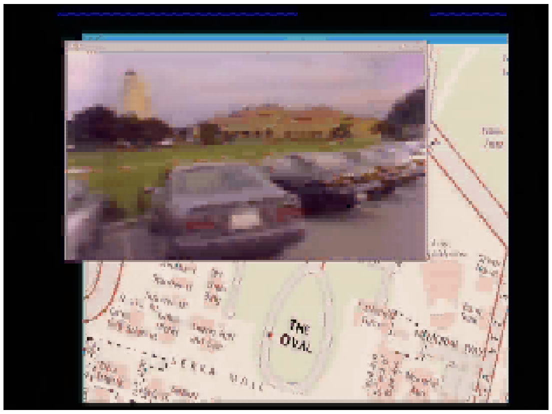

Figure 8: *This image shows a single frame from the video panorama project. The video panorama consists of a tour of Stanford campus. At each point in time a full 360 degree panorama is received and the user is free to browse around, but a default view is provided. The display also shows a map of the campus with a pointer indicating the current position. By clicking on a location on the map the current viewing direction rotates and locks onto this feature allowing a user to watch it while the tour continues.*

MPEG-2 decoder software has been modified so that eight copies can be run synchronously delivering frames in the YUV format. A master process controls the eight decoders, and requests the frames that are needed to construct the current view. These frames are then passed on to the panorama viewer. We have chosen to always decode all eight video streams due to the startup delay of about half a seconds when decoding MPEG-2 streams with large GOPs (group of pictures). This delay would have severely restricted the speed at which viewers could have changed their view.

The viewer program is an OpenGL application that texture maps the video frames on to eight rectangular polygons which form an overlapping octagon around the virtual camera (see Figure 7). The spherical aberration of the capture cameras is corrected by suitable warping of the input video on to each display polygon. During display, the viewer application determines the frames that are needed based on the current view direction and requests them from the MPEG-2 decoder system. The images are loaded as OpenGL textures, and a color transformation matrix is used to convert them from YUV format into RGB in hardware. For a given camera rig, the exact arrangement of the display polygons needs to be adjusted so that the frames appear stitched together into a seamless panorama. Currently, this is accomplished by manually adjusting their positions, but automatic calibration is planned for a future version of the application.

In addition to the basic panorama viewer, several additional features were added to make the video panorama more useful (see Figure 8 and the video clip). Since the captured material was a tour of Stanford campus, a campus map is displayed in addition to the video panorama window. This map displays the position and direction of the current view. The user can point on the map to indicate a new viewing direction, e.g. toward a certain object, which is then tracked. This is useful for getting ones bearings as the tour progresses. Finally, on command or in cases when the user stops interactions, the view drifts back towards a predefined view sequence chosen by the tour guide. Thus, lazy viewers are guaranteed to be looking in the direction of the object currently being discussed by the guide on the associated audio track. At any time, the user has the freedom to override the tour guide or again join his tour, very much like in reality. Combining the current application with the Virtual VCR support (see Section 5.2) would also allow a user to stay behind and enjoy a particular view before catching up or resuming the tour. This is planned as a future extension.

6. Conclusions

In this paper we explored the emerging digital television technology and its interaction with computer graphics. DTV is supposed to provide high bandwidth broadcast connections to every household and offers high-quality, high-resolution video streams. Most interesting from a computer graphics perspective is the ability to transmit any other data with the DTV stream, which allows us to extend DTV in a number of different directions.

In the previous sections we presented several applications as examples from the large DTV design space. We provided the ability to decouple the display time from the transmission time through a real-time *video buffer*, which is an important base technology that we applied to a number of the other applications. We showed how to provide additional information onto the live video screen in the form of *video overlays*. The overlays are manipulated by small scripts that control their appearance and allow the user to interactively browse their information. In the *lecture of the future* project, we concentrated on using image-based techniques to scan and combine small high-resolution images of a blackboard and transmit them as a separate video stream to enhance the key lecture part. Finally, we presented a virtual reality application, using real-time *video panoramas* that provide new visual experiences and interactivity in the context of DTV transmissions.

In conclusion, we hope to have demonstrated that computer graphics offers considerable know-how and technology in areas such as image-based rendering, interactive applications, and virtual reality, that can directly contribute to creating exciting new DTV applications. As digital television technology becomes more widely available over the next few month and years, we expect to see many new developments that may forever change our notion of television.

Acknowledgments

First of all, we would like to thank Pat Hanrahan and Serge Rutman (INTEL) for sharing their experience. Many thanks are also due to all the students who spent many hours in designing and implementing the presented projects. In particular we would like to thank Edward Chang, Ben Mowery, Hareesh Kessavan, Brad Nelson, Chuck Fraleigh, Rustan Eklund, Bryn Forbes, and Jin Hian Lee for providing direct input to this paper. We also want to thank Tamara Munzer for helping with the accompanying video. Finally, we like to thank Intel, Sony, and Interval Research for funding the Stanford Immersive Television Project, which includes most of this work.

References

1. Albert Abramson. *The History of Television, 1880 to 1941*. McFarland, 1987.
2. ATSC homepage. http://www.atsc.org/.
3. Edward Chang and Hector Garcia-Molina. Effective memory use in a media server. In *Proceedings of the 23rd Very Large Data Base (VLDB) Conference*, pages 496–505, Athens, Greece, August 1997.
4. Shenchang Eric Chen. Quicktime VR - an image-based approach to virtual environment navigation. *Computer Graphics (SIGGRAPH '95 Proceedings)*, pages 29–38, 1995.
5. Shenchang Eric Chen and Lance Williams. View interpolation for image synthesis. *Computer Graphics (SIGGRAPH '93 Proceedings)*,, 27:279–288, August 1993.
6. Carolina Cruz-Neira, Daniel J. Sandin, and Thomas A. DeFanti. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. *Computer Graphics (SIGGRAPH '93 Proceedings)*, pages 135–142, 1993.
7. Gary Demos. Temporal and resolution layering in advanced television. http://home.earthlink.net/ demografx/layered.html, 1995.
8. DVB homepage. http://www.dvb.org/.
9. David E. Fischer and Marshall Jon Fisher. *The Tube – The Invention of Television*. Harcourt Brace, 1996.
10. Barry G. Haskell, Atul Puri, and Arun N. Netravali. *Digtal Video – An Introduction to MPEG-2*. Chapman & Hall, 1997.
11. Nelson L. Max. Computer graphics distortion for IMAX and OMNIMAX projection. *Nicograph '83 Proceedings*, pages 137–159, December 1983.
12. MPEG audio. http://www.mpeg.org/MPEG/audio.html.
13. MPEG-2 – generic coding of moving pictures and associated audio information. online available at http://drogo.cselt.stet.it/mpeg/standards/mpeg-2/mpeg-2.htm, July 1996. ISO/IEC 13818.
14. MPEG-4 standard, version 1. available at http://drogo.cselt.stet.it/mpeg/, October 1998.
15. Stephan Oettermann. *The Panorama – History of a Mass Medium*. Zone Books, 1997.
16. Charles A. Poynton. *A Technical Introduction to Digital Video*. Wiley, 1996.
17. Charlie Sandbank and Ken McCann. DVB and HDTV. In *Proceedings IBC'97*, Sydney, Australia, December 1997. online available at http://www.dvb.org/dvb_articles/dvb_hdtv.htm.
18. Jonathan W. Shade, Steven J. Gortler, Li wei He, and Richard Szeliski. Layered depth images. *Computer Graphics (SIGGRAPH 98 Conference Proceedings)*, pages 231–242, July 1998.
19. Alvy Ray Smith. Digital TV postgame wrapup. http://www.research.microsoft.com/Alvy/DigitalTV/NAB%20Postgame%20Wrapup.htm, August 1998.