THE CHALLENGE OF DVD AUTHORING

Panos Nasiopoulos[1], Rabab K. Ward[1] and Masato Otsuka[2]
[1] Electrical Engineering, University of British Columbia, Vancouver, BC, Canada
panosn@ee.ubc.ca   rababw@ee.ubc.ca
[2] DAIKIN US Comtec Laboratories, Novato, CA, USA
masato@daikin-comtec.com

Abstract: The evolution of DVD promises telecomputers will find their way into our living rooms, linked to a large flat screen which will be able to display HDTV, Standard Definition TV, video games, interactive movies from Digital Versatile Discs (DVD), Internet, video telephony and computer graphics. For the first time, Hollywood Studios and consumer electronic companies have formed an alliance to support the DVD technology which is critical for both sides. In this paper we address the complex process of DVD authoring.

1. INTRODUCTION

One of the most significant technological achievements in the consumer and entertainment industries is the development of the Digital Versatile Disc (DVD). DVD is the first union of emerging technologies, bringing together computer consumer electronics and entertainment. As a result, we are witnessing the generation of an entirely new infrastructure that is reshaping the world of entertainment.

DVD is a lot more than just a storage medium. It is a new multi-purpose technology that will affect both the entertainment and computer worlds. For consumers, this is the first digital medium that offers studio video and audio quality combined with unprecedented interactivity at a very low cost. For the PC multimedia side, DVD is the first video distribution medium designed for very high data rates. Its interactivity and distribution format have the potential to revolutionize the entertainment software industry. This paper describes the complex process of DVD authoring.

2. DVD: AN OVERVIEW

2.1. Storage

As a storage media, DVD can hold from 4.7 GB of digital data on one-side single-layer format to 17 GB on a double-side dual-layer format [1,2]. This increase in capacity (up to 25 times that of CD's) is achieved by introducing a shorter wavelength laser beam, dual focusing mechanism that allows the use of two layers per side, smaller pit size and tighter spirals. Furthermore, DVD discs offer ten times the speed of the CD rate, opening the way to numerous new real-time applications. While storage capacity is very important, it is DVD's other capabilities that make this technology so attractive.

2.2. Hollywood

As an entertainment product, DVD satisfies the goals established by the Hollywood Digital Video Disc Advisory Committee, delivering extraordinary picture and sound quality. DVD takes advantage of a two-pass variable bit rate MPEG-2 video encoding process to offer a superb picture quality comparable to D-1, the studio production standard. To make it more exciting, this is the first medium to introduce a number of viewing formats such as the 4:3 TV screen format, the 16:9 HDTV screen format and the 20:9 letterbox format [1,2]. Combine this picture experience with Dolby's AC-3 5.1 channel surround sound (or MPEG-2 7.1 for Europe) and you have reproduced video and audio quality that rivals that of a theater.

In addition, this technology allows the use of 8 different languages, 32 subtitles, different camera angles and video-clip paths including interviews with producers and actors. The viewer can choose the camera angles and language, switching seamlessly from one to another, scan forwards and backwards and play slow motion. Parents are given the option to lock out versions of the movie which range from the director's cut, to R-rated, to PG-13.

2.3. Hybrid DVD-Internet

But it is the PC world where DVD will have its biggest impact. The first read-only DVD drives are expected to offer over 7 times the storage capacity of the current CD, and will also be able to play DVD movie titles and existing music CDs. DVD brings the added capability of supporting the implementation of interactive adjuncts to traditional PC content. Embedded navigation such as web browsing can be added to video, enhancing the user's experience with hybrid content.

A DVD-based PC application that combines DVD's interactive performance and rich video and audio capabilities with the Internet would offer a wealth of opportunities. For example, it is possible to produce a DVD-based department store catalogue that offers an interactive showcase of all the departments' merchandise, complete with audio and video. Playing the disc automatically connects the user with the store via the Internet, allowing the consumer to get current prices,

order merchandise, communicate with a personal shopper or pay bills. Such a service is not viable today because of the Internet's low bandwidth. A similar hybrid application allows DVD-based courses and encyclopedias (which devour huge amounts of space with text, pictures, video sound and animations) to remain up-to-date by cross-referencing a constantly updated web site.

3. DVD AUTHORING

Authoring is generally defined as the process of preparing content, encoding video and audio, and creating the final DVD image. In the case of DVD, authoring is a complex process since it involves the laying out of multiple audio tracks and a video track, generation of sub-titles, menu pages, parental lock-out features, interactive functions such as program search, time search, seamless play, and pause, and finally editing of video and audio. Since authoring is always performed along with encoding and disc formatting, it is, in many cases, referred to as the entire DVD pre-mastering process (see Figure 1).

3.1. Preparation of Materials

The first step in authoring is the collection of materials. These materials include video, audio, still images, and sub-pictures. DVD's video source format is the CCIR-601 studio format compressed to MPEG-2 format. The frame rate is 29.97 f/s for NTSC sources (North America) and 25 f/s for PAL/SECAM sources (Europe). The maximum allowable bit rate is 9.8 Mbps. Audio includes the surround track and up to 8 different language tracks for each title. All language tracks must be compared for level, mix, and equalization so that seamless switching between languages can be achieved. Still images are used to provide break points in the title, so that search functions and other interactive functions can be implemented. The preparation of still images includes identification of the breakpoints in the video and definition of the time duration for each image. Sub-pictures are bitmaps that are overlaid on top of the video. They include menus, sub-titles, graphics, and simple animation. Once created, their start and stop time must be defined in order to be synchronized with the associated video and audio elements. Up to a maximum of 32 sub-picture bit-streams are allowed in a title.

3.2. Techniques and Parameters

Good understanding of how various elements will be used in constructing the title is the key to intelligent parameter determination which includes tradeoffs between picture quality, length of program, number and quality of audio channels, number of subtitles, and level of interactivity. The following is a list of some of the basic parameters needed to be determined for a DVD title [3]:

- the number of audio channels
- the number of language versions
- the number of sub-picture elements
- the number of breakpoints in the video
- the number and the levels of rated versions of the title
- the number of still images used at each breakpoint
- the type of parental lock outs
- the type of directors cuts
- the audio encoding techniques
- the format used for still images

A single-layer single-sided DVD disc can store 2 hours and 13 minutes of video compressed at a nominal average bit rate of 3.5 Mbps combined with 3 languages encoded using AC-3 5.1 channels and 4 additional languages encoded as sub-titles [4]. The maximum program rate (i.e., video + audio + sub-pictures) is specified to be 10.08 Mbps. Given the disc capacity, the overall quality depends on determining the different trade-offs between several parameters. For example, Table 1 shows the average storage requirements for a DVD title with the following parameters:

- Audio tracks encoded using Dolby AC-3 5.1
- 4 unique languages supported
- 4 sub-picture streams supported
- "G" rated version has a total run length of 100 minutes
- "PG" rated version has an additional 4 minute run length
- 2 previews; each has a run length of 3 minutes
- 4 trailers; each has a run length of 2.5 minutes

Note that 4% of the total disc capacity is always reserved for backup of the program control data and for additional information that is added after editing. The total run length is 120 minutes, resulting in the average bit rate of 3.43 Mbps.

3.3. Video Encoding

DVD takes advantage of the MPEG-2 compression technology to achieve picture quality comparable to that of D-1, the CCIR-601 TV studio production standard. MPEG-2 is a flexible and scaleable compression scheme which can produce bit rates that range from 1 to 40 Mbps. As implemented for DVD, MPEG-2 encoding is a two-pass process. During the first pass, the encoder scans the video source, detects scene changes and determines the optimal bit rates for each frame. During the second pass, higher bit rates are assigned to complex frames and sequences with more activity and lower bit rates to "simple" frames. The two-pass process guarantees the best possible picture quality for the given video clip and disc storage capacity.

For video material originated from film, inverse telecine may be used to improve the compression performance. The reason is that film uses 24 f/s, a rate that is converted to the 30 f/s required by the NTSC standard. This conversion process is known as telecine and involves duplication of frames at regular intervals. Inverse telecine removes the duplicated frames, thus allowing more bandwidth to be allocated to the video. Figure 2 illustrates the two-pass video encoding process.

3.4. Audio Encoding

Movies released in North America and Japan can carry Dolby's AC-3 stereo or 5.1 audio which offers 5 surround channels plus a low frequency (sub-woofer) channel. For movies released in Europe, AC-3 is replaced by MPEG-2 stereo or 7.1 surround sound. In addition, as an option to AC-3 and MPEG-2 audio, DVD enables producers to choose uncompressed 16-bit linear PCM stereo sound with Dolby Pro Logic encoding. Table 2 shows audio encoding options as well as the specified sampling frequency rates, bit and transfer rates and number of channels supported by each option.

3.5. Sub-Picture Encoding

Sub-pictures are run-length compressed bit-maps using 2 bits/pixel and 4 colors out of a 16 color palette. The sub-picture size is 62KB per GOP/cell with 32 KB allocated for control data. Applications may vary from simple text (sub-titles) to menus to still images used for presentation effects. Pixels are categorized as foreground, background, emphasis-1 and emphasis-2. The still picture format must be a standard image format such as TIFF, GIF, or BMP. MPEG is used to encode still images which are then incorporated into the video stream.

3.6. Putting it Together

After preparing the different "segments" of a DVD title, a multiplexing process should link everything together and define the program flow of the DVD title. This final step should specify how each of the media elements will be presented to the user and how the user can interact with the program. Program flow specifications are translated to navigation commands that are, in turn, incorporated into program cells and program chains. A cell consists of a navigation command and all the video and audio data associated with a GOP (see Figure 3). The navigation command (button) defines the playback behavior of the corresponding cell and it consists of one or at most a combination of three of the following instructions [4]:

• GoTo	→	branch between commands
• Link	→	transfer between the same domain
• Jump	→	transfer between each domain
• Compare	→	recognition of parameter value
• SetSystem	→	player system setting
• Set	→	calculate GPRM values

A sequence of cells and cell commands (navigation commands) form a program (PG). A program usually corresponds to one scene. Programs and video objects (nominally a GOP) form a program chain (PGC). A program chain is separated into the control information (PGCI) and the video object (VOB) as shown in Figure 4. PGCI acts as an address table pointing to cells, thus defining the playback order of Programs.

The Part of Title (PTT) helps to construct multiple versions of the same title. A DVD title can have only one or multiple program chains (see Figure 5). Interactive functions such as PTT searches, director's cuts, and parental lock-outs can be achieved by creating the title as a multi-PGC_title, with different director's cuts and different rated versions on different program chains.

3.7. Simulation and Verification

After all the media elements and control information are multiplexed into one stream, simulation testing is to be performed. The stream must guarantee that audio, video, and sub-pictures are synchronized; otherwise, the content must be re-edited or re-encoded. Besides synchronization, interactive functions may also be simulated and verified.

4. REFERENCES

[1] DVD Format, TOSHIBA, DVD Forum April 1996.
[2] DVD Presentation Data Specifications, VICTOR Company of Japan Ltd., DVD Forum, April 1996.
[3] C. Fogg, DVD Technical Notes, July 1996.
[4] Interactive Functions, HITACHI Ltd., DVD Forum, April 1996.

Table 1. Storage Requirements for Each Media Element in Average-Bit-Rate Calculation Example

Media Element	Total Length	Average Bit Rate	Total Storage Requirements
4 Language Tracks	120 minutes	0.384 Mbps per language	4*120*60*0.384Mbps/8 = 1382 MB
4 Sub-picture streams	120 minutes	0.01 Mbps per language	4*120*60*0.01Mbps/8 = 36 MB
Reserved		4% of 4.7 Gbytes	188 MB
		SUBTOTAL	1606 MB
Video	120 minutes	3094/(120*60)*8= 3.43 Mbps	3094 MB

Table 2. Audio Data Specifications

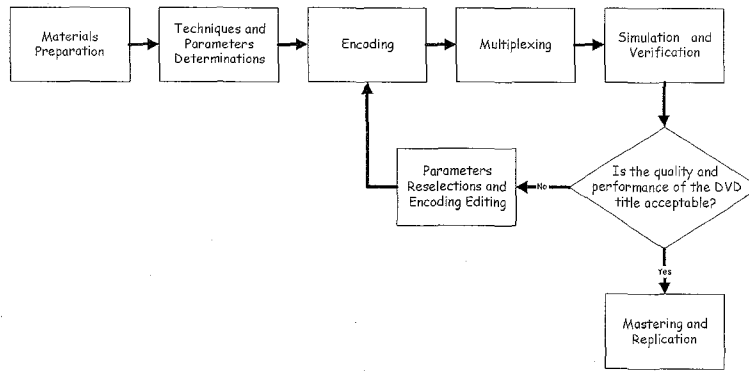|	Linear PCM	Dolby AC-3	MPEG Audio
Sampling Frequency	48K, 96K	48K	48K
Number of Bits	16/20/24 bits	compressed	compressed
Transfer Rate	max. 6.144 Mpbs	max. 448 kbps	max. 640 kbps
Number of Channels	max. 8	max. 5.1	max. 7.1

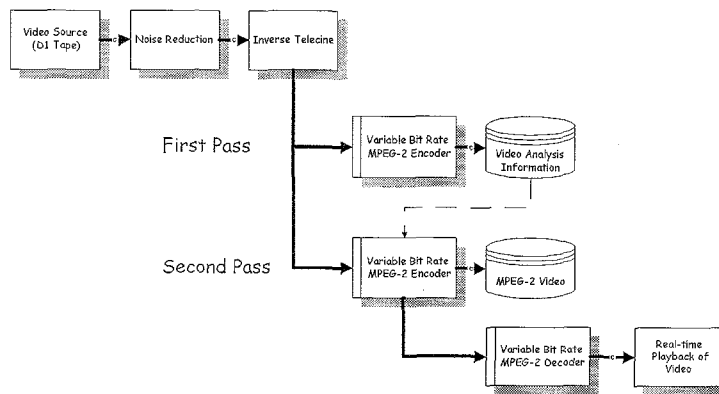Figure 1. The Pre-mastering Process Flow

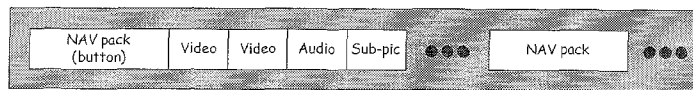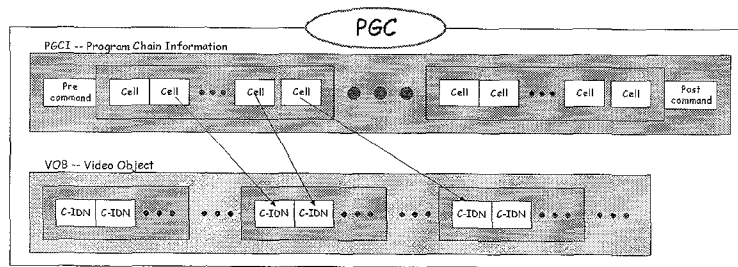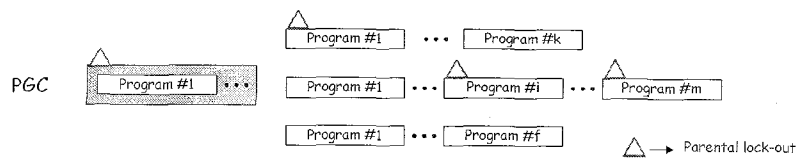Figure 2. Two Pass MPEG-2 Video Encoding Process

Figure 3. A Program Cell

Figure 4. Program Chain Structure

Figure 5. A Multi-PGC-Title