

DVD-Video 구현을 위한 VOB Encoding에 관한 연구

장형일 • 박인규

홍익대학교 내학빈 전기제어공학과
121-791 서울 마포구 상수동 72-1
garfield@wow.hongik.ac.kr

A Study on VOB Encoding for Implementing DVD-Video System

Hyung-il Jang • In-gyu Park

Dept. of Electrical & Control Eng.,
Graduate School, Hong Ik University
72-1 Sangsu-dong Mapo-gu Seoul 121-791
garfield@wow.hongik.ac.kr

요약

비디오 스트림과 오디오 스트림 그리고 제어신호등의 서로 다른 종류의 digital data들을 저장 또는 전송하기 위해서는 하나의 스트림으로 통합하기 위해 다중화 과정을 거쳐야 한다. 이러한 다중화 과정에서 가장 중요한 문제점은 하나로 묶여진 일련의 스트림을 원래의 순차대로 복원해주는 동기화 문제이고, DVD-Video에서는 MPEG2 system standard인 ISO/IEC 13818-1에서 제시한 time stamp를 이용한 동기방식을 사용한다.

본 논문은 DVD 규격에 맞추어 Microsoft의 Visual C++5.0을 이용하여 software로 VOB를 구현하였다. 입력 스트림은 MPEG2로 압축된 비디오스트림과 최대 8개의 5.1채널 AC3 오디오스트림이고 각 스트림은 frame단위로 구성이 된다. 이 frame들을 단위로 time stamp가 부여되며 time stamp의 비교에 의해 우선 순위를 부여하여 encoding을 하게된다. 그리고 같은 time stamp를 가졌을 경우 발생하는 encoding 순서의 문제는 스트림별로 미리 우선 순위를 정함으로써 해결했다.

1. 서론

차세대 영상 매체로 불리는 DVD(Digital Video Disk)는 원래 미국 영화업계의 요구 때문에 규격 개발이 시작되었다. 고화질의 영상과 음질을 보장하면서 2시간이 넘는 시간을 재생하기엔 기존의 Video CD로서는 너무나 부족했기 때문이다. DVD는 단층구조의 경우 4.7GByte, 이중구조의 경우 8.5GByte의 저장용량을 갖음으로서 기존 CD에 비해 7-13배 이상의 저장용량을 갖는다.

일반적으로 영상을 디지털화 하면 데이터 전송률이 초당 20MB정도이고, 131분간의 영상과 음질을 담으려면

CD의 250배나 되는 용량의 디스크가 필요하다. 이 때문에 영상 및 음성 압축 기술이 필요한 것이다. DVD-Video에서는 영상 압축 표준으로는 MPEG2(Moving Picture Expert Group 2)를, 오디오의 경우는 돌비사의 5.1채널 AC3(Audio Compression3)을 채택하고 있다.

MPEG2는 고화질 TV등에 사용되는 비디오 데이터 압축 방법으로 이 규격을 사용한 DVD의 화질은 LD(Laser Disk)수준을 능가한다. 즉 극장에서 보는 영화와 흡사한 고화질의 영상을 제공한다. 또한 DVD에서 사용하는 AC3기술은 단순히 스테레오가 아니라 5.1채널로 소리를 들려준다. 이것은 앞쪽 스피커 3개, 뒤쪽 스피커 2개 그리고 저음용 강조하는 0.1채널을 일컫는 말로 모든 스피커에서 다른 소리가 나오므로 극장을 능가하는 System이다.

그리고 최대 8개 국어로 더빙되어 다국어틀 지원하며, 최대 9개의 다른 영상들이 여러 대의 카메라에 의해 동시에 기록될 수 있기 때문에 보는 사람의 선택에 의해 원하는 각도에서 시청을 할 수 있다.

이렇게 멀티미디어의 발전이 가져다 준 DVD-Video의 등장으로 현재 TV와 오디오, 컴퓨터로 분리되어 있던 전자제품군이 DVD를 축으로 하나로 통합될 것으로 보이며 이에 DVD-Video의 Stream인 VOB의 Encoding System이 절실히 요구된다.

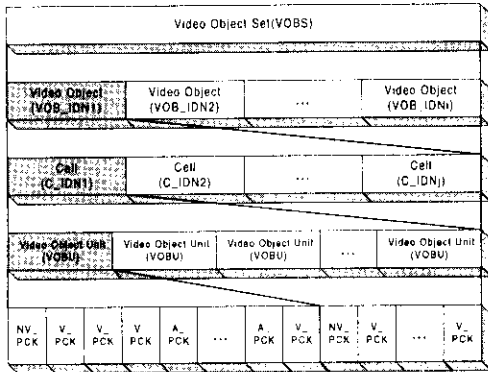
본 논문은 구성은 다음과 같다.

2장에서 VOB의 구조와 Navigation data내에 알아보고 3장에서는 동기화를 위한 알고리즘에 대해 설명한다. 4장에서는 VOB를 구현한 소프트웨어에 대해 설명하고 마지막 5장에서는 결론 및 향후 계획에 대해서 기술한다.

2. DVD-Video의 Logical Data Structure

1. VOB

VOB는 Video data, Audio data, PCI data 및 DSI data 로 구성되며 그림 1에서 보는 바와 같이 여러 하위구조 들로 분류를 할 수 있다.



<그림 1 : VOB의 구조>

VOB(Video Object)는 Multi-Angle 및 Multi-Story를 지원하기 위한 Cell들로 분류가 되고, Cell은 VOB라고 하는 MPEG 비디오에서의 GOP개념에 해당하는 Access의 최소단위로 구성이 된다. VOB의 Play time은 GOP의 Play time과 같으며, GOP Play time에 해당하는 오디오 데이터와 PCI와 DSI가 Navigation pack이란 하나의 pack의 형태로 VOB의 첫 번째 팩으로 포함된다.

VOBU를 구성하는 Pack들은 모두 2KByte의 크기를 가지며 Navigation pack다음 pack은 항상 Video pack이 오며 VOB의 마지막 팩 역시 Video pack이 된다. 이것은 모두 VOB가 MPEG의 GOP를 기준으로 구성이 되기 때문이다.

Pack은 pack header에 하나의 PES가 결합된 형태로 구성이 된다. PES를 구성하기 위해서 packet header를 제외한 elementary stream이 pack에서 차지하는 크기를 알아야 한다. Pack header는 14바이트로 모든 팩에서 똑같은 크기를 가진다. 그러나 packet이 access unit을 포함하느냐에 따라 PTS와 DTS의 포함여부가 결정이 되므로 packet header의 길이는 달라진다.

(1) MPEG-2 Video Pack의 경우

① MPEG-2 Video stream에서 GOP의 처음부분 즉 sequence_header_code를 포함하는 경우:PTS와 DTS존재

$$\begin{aligned} \text{Packet Size} &= 2048(\text{Pack Size}) - 14(\text{Pack Header}) \\ &\quad - 22(\text{Packet Header}) \\ &= 2012 \text{ Bytes} \end{aligned}$$

② sequence_header_code를 포함하지 않는 경우:Stamp 존재 안함

$$\begin{aligned} \text{Packet Size} &= 2048(\text{Pack Size}) - 14(\text{Pack Header}) \\ &\quad - 9(\text{Packet Header}) \\ &= 2025 \text{ Bytes} \end{aligned}$$

(2) AC-3 Audio Pack의 경우

① AC3 Audio stream에서 Frame Synch(0x0B77)를 포함하는 경우 : PTS만 존재

$$\begin{aligned} \text{Packet Size} &= 2048(\text{Pack Size}) - 14(\text{Pack Header}) \\ &\quad - 17(\text{Packet Header}) \\ &= 2017 \text{ Bytes} \end{aligned}$$

② Frame Synch를 포함하지 않는 경우:Stamp 존재 안함

$$\begin{aligned} \text{Packet Size} &= 2048(\text{Pack Size}) - 14(\text{Pack Header}) \\ &\quad - 9(\text{Packet Header}) \\ &= 2025 \text{ Bytes} \end{aligned}$$

2. Navigation Data

Navigation data에는 VOB의 재생에 필요한 정보를 포함한 PCI Packet과 Data Search 및 Seamless 재생, 동기화 정보가 포함된 DSI packet으로 구성이 된다.

(1) PCI(Presentation Control Information)

	content	Number of bytes
PCI_GI	PCI General Information	60
NSML_AGLI	Angle Information for non-seamless	36
HLI	highlight Information	694
reserved	reserved	189
	Total	979

<표 1 : PCI Packet>

(2) DSI(Data Search Information)

	content	Number of bytes
DSI_GI	DSI General Information	32
SML_PBI	Seamless Playback Information	148
SML_AGLI	Angle Information for Seamless	54
VOBU_SRI	VOBU Search Information	168
SYNCI	Synchronous Information	144
reserved	reserved	471
	Total	1017

<표 2 : DSI Packet>

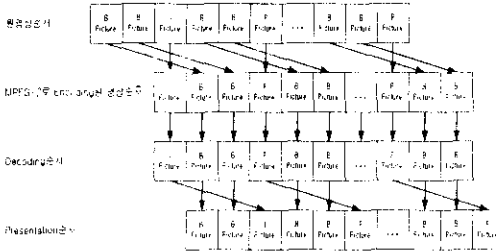
3. 동기화를 위한 알고리즘

(1) time stamp를 이용한 동기방식

비디오, 오디오를 각각 packet으로 분할하고 packet header와 pack header에 동기화 관련된 time stamp와 SCR을 포함시켜 시분할 전송하고 decoder측에서는 decoder의 STC(System Timing Clock)가 PTS나 DTS와 일치할 때 재생 또는 decoding하는 방식을 사용하여 동기화를 이룬다.

SCR(System Clock Reference)는 decoder의 기준 시각 정보인 STC(System Timing Clock)을 encoder측에서 의도한 값으로 재 설정하기 위한 값으로 PLL을 고려하여

27MHz를 clock을 기준으로 측정한 값을 사용한다.



<그림 2 : 영상의 표시 순서>

위그림은 MPEG으로 압축을 하기 전과 압축 후의 영상의 표시 순서를 나타낸 것이다. decoder측에서 decoding은 입력되는 영상의 순서대로 이루어진다. 위 그림에서 Decoding 되는 순서에 따라 Stamp를 부여하면 stamp의 순서는 (I₀, B₁, B₂, P₃, B₄, B₅...)순서로 부여될 것이다. 그러나 Presentation 순서는 (B₁, B₂, I₀, B₁, B₃, P₄...)로 뒤바뀌게 된다. 즉 먼저 재생해야 하는 picture가 나중에 decoding 되는 현상이 발생한다. 그래서 MPEG2 비디오의 경우는 Presentation 순서를 나타내는 PTS와 Decoding 순서를 나타내는 DTS의 두 가지 time stamp가 존재한다.

반면 오디오의 경우는 decoding 순서가 presentation 순서와 동일하기 때문에 PTS 한가지 stamp만이 존재한다.

PTS와 DTS의 정밀도는 90KHz의 clock으로 측정된 값을 33비트로 나타냄으로써 하루 24시간을 모두 표현할 수 있다.

(2) packet의 우선순위

각 비디오와 오디오 스트림은 개별적으로 부여된 time stamp가 부여된다. 이 값은 매 프레임이 encoding 될 때마다 한 프레임 재생시간 만큼 더해진 값으로 update가 된다.

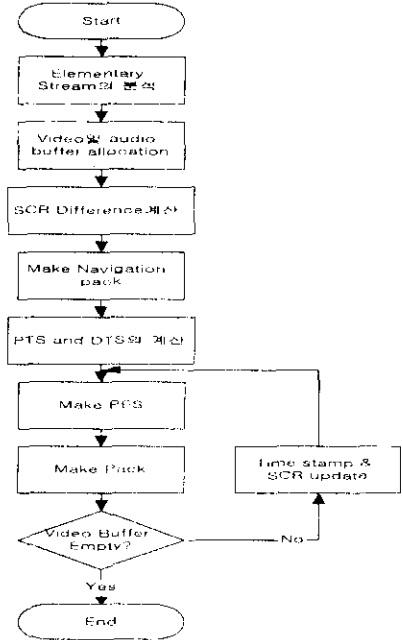
이 time stamp들의 비교를 통해 가장 작은 time stamp를 가진 stream을 선정하여 해당 stream에 우선 순위를 부여하고 packetize를 통해 PES로 만들어진 후 pack으로 encoding된다 한 프레임의 Encoding이 진행이 되면 다음 Frame이 나올 때까지는 우선권을 계속 가지고 있다.

그리고 이 time stamp를 비교했을 때, 같은 값을 가질 경우는 다음과 같은 우선 순위를 부여한다.

Video>Audio0>Audio2>...>Audio7

4. 소프트웨어의 구현

아래 그림은 하나의 VOB를 구성하는 소프트웨어의 순서도를 나타낸 것이다.



<그림 3 : software 순서도>

(1) Elementary stream분석

입력되는 비디오 스트림 즉 MPEG-2 비디오를 Offline으로 scan하여, 실제 SCR계산, PTS계산 또 실제 다중화시 필요한 정보들과 GOP의 크기를 GOP단위로 Linked list에 저장한다.

(2) Video & Audio buffer allocation

빠른 Encoding을 하기 위해 버퍼를 할당하여 encoding을 한다.

MPEG2 비디오의 경우는 한 GOP크기만큼 buffer를 allocation한다. GOP의 크기는 달라지므로 VOB나 buffer의 크기는 변화하게 된다. 버퍼의 크기는 해당 GOP의 Linked list를 검색하여 알 수 있다.

오디오 버퍼의 크기는 4KByte로 할당해주며 이는 매 오디오 팩이 encoding될 때마다 update된다.

(3) SCR Difference 계산

$$SCR\ Difference = \left(\frac{System\ Clock\ Frequency}{Program\ mux\ rate} \right) \times Sector\ size$$

여기서 System Clock Frequency는 27MHz로 정의되며 Program Mux Rate는 10.08Mbps로 1,260,000Mbps(Byte per sec)가 된다. 그러면 한 바이트 전송시 걸리는 시간은 $\frac{1}{1,260,000} Sec$ 가 되고 한 Sector의 크기는 2KByte = 2048Byte이므로 한 Pack 전송 시에 필요한 시간

$$\text{Sec Per Pack transmit} = \frac{2K\text{Byte}}{1,260,000 \text{ Byte Per Sec}} \text{ 가}$$

$$= 0.001625\text{Sec}$$

된다.
이것을 27MHz로 측정한 값으로 환산하면
 $0.001625 \times 27\text{MHz} = 43875$ 가 된다. 이것이 SCR Difference e이다.

VOBU의 최초의 팩의 SCR은 반드시 0이어야 하며, 두 번째 팩부터는 SCR Difference 만큼씩 더해줌으로써 SCR을 구할 수 있다. 그런데 이 SCR의 값이 Pack Header에 적용될 때는 이 구해진 SCR을 300으로 나눴을 때의 몫인 SCR Base와 나머지인 SCR Ext로 분리하여 Marker bit와 함께 헤더에 포함 시켜야 한다. 이 크기는 SCR Base+SCR Ext=12bit와 Marker bit=4비트로 46비트이며 Pack start code 다음에 나타나는 "01" 즉 MPEG-2임을 나타내는 비트와 함께 6바이트를 이룬다.

(4) Make Navigation Pack

VOBU의 최초의 팩이다. PCI packet과 DSI packet으로 구성하고, System header를 가진다.

이 System header에는 비디오 및 오디오 bound, 비피의 크기 및 PES packet에 대한 내용을 기술한다.

(5) PTS 및 DTS의 계산

① Time per frame의 계산

a. MPEG2 비디오

MPEG2 비디오에서 Frame rate code를 나타내는 부분은 4비트의 값으로 다음과 같이 MPEG2의 Header에 포함되어 있다.

Sequence Header Code	Horizontal Size Value	Vertical Size Value	Aspect Ratio Information	Frame Rate Code	Bit Rate Value	Marker Bit
32bits	12bits	12bits	4bits	4bits	18bits	1bit
0000103h				동영상 표시기		동영상 표시기

<그림 4 : MPEG-2 Stream Header>

picture_rate[9]={0,24000/1001,24,25,30000/1001,30,5,0,60000/1001,60}

```
if(Frame_Rate_Code>0 && Frame_Rate_Code<9)
    sec_per_frame=
        (float)(1./picture_rates[Frame_Rate_Code])
else
    sec_per_frame=(float)(1./25.);
```

$$\text{time per frame} = \text{sec per frame} \times 90,000$$

$$= \frac{1}{\text{picture rates[Frame Rate Code]}} \times 90,000$$

b. AC3 오디오

Sync	CRC	fscode	Frmsize
16Bits(2Bytes)	16Bits(2Bytes)	2Bits	6Bits

<그림 5 : AC3 Stream header>

fscode와 Frmsize code를 있어서 Frame size와 Bit rate를 구하고 이를 바탕으로 time per frame을 구한다.

$$\text{time per frame} = \left(\frac{\text{Frame size}}{\text{Bit rate}} \right) \times 90,000$$

② time stamp의 초기화

처음으로 재생이 되는 시간을 초기화 한다. 90kHz로 clock 기준으로 6006으로 초기화한다. 이는 29.97 Frame/sec경우 두 화면에 해당하는 시간이다.

$$\text{Initial time stamp} = 6006$$

③ DTS 및 PTS 계산

$$\text{Initial Audio PTS} = \text{Initial Time Stamp}$$

$$\text{Initial Video DTS} = \text{Initial Time Stamp} - \text{time per frame}$$

$$\text{Initial Video PTS} = \text{Initial Video DTS}$$

$$+ (\text{time per frame} \times \text{first temporal ref})$$

next PTS와 DTS값은 frame이 하나마다 time per frame을 더해줌으로써 구할 수 있다

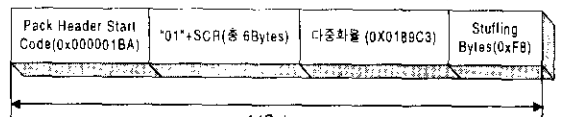
(6) Make PES

Audio PTS와 (Video DTS+time per frame)을 비교해서 낮은 쪽에 우선 순위를 부여하여 우선 순위를 가진 stream을 packetize하여 packet header를 붙여 PES를 만든다.

(7) Make Pack

PES에 pack header를 붙임으로써 pack을 만든다.

pack header의 크기는 14바이트로 decoder측의 clock을 조정하기 위한 SCR과 다중화율이 표시된다.



<그림 6 : Pack Header>

5. 결론 및 향후계획

본 연구에서는 DVD규격 맞추어 DVD-Video를 위한 VOB를 소프트웨어로 구현하였다. 규격에서 제시되지 않은 각 스트림들의 encoding순서는 time stamp의 비교를 거쳐 스트림별로 우선 순위를 부여함으로써 해결하였다. 차세대 디지털 방송에 쓰일 HDTV의 규격 또한 MPEG2와 AC3를 표준으로 채택하고 있다. 본 연구에서는 DVD를 target으로 구현했기 때문에 실 시간성을 고려하지 않았으나 앞으로 별도 하드웨어의 추가로 다중화가 실시간으로 처리되면, 실시간 MPEG2 encoding system, 실시간 AC3 encoding system과 함께 HDTV를 위한 실시간 디지털 방송장비를 구현할 수 있다.

[참고문헌]

1. DVD Specification for Read-only Disk Ver0.9, 1996
2. ISO/IEC. Generic coding of moving pictures and associate audio information Part1: systems Part2: Video. CD 13818. May 1995
3. ATSC. Digital Audio Compression Standard (A52). 1995