

IP 방송제작 기술 9

비디오 데이터 전송

글. 조인준 KBS 미디어기술연구소 차장

이전의 IP 방송제작 기술 시리즈를 통해 PTP 동기화 및 PTP를 이용한 타임코드 발생에 대해 다루어왔습니다. 이번 연재부터는 IP 방송제작 기술에서 가장 중요하다고 할 수 있는 비디오 데이터 전송에 대해서 알아보겠습니다. 현재 상용화된 IP 방송제작 장비에서 사용하는 비디오 데이터 전송 방식은 몇 가지가 존재하지만, 본 시리즈에서는 국제공개표준인 SMPTE 표준의 IP 비디오 데이터 전송에 대해 다루도록 하겠습니다.

SMPTE에서 제공하는 IP 비디오 데이터 전송 관련 표준은 SMPTE 2022 계열과 SMPTE 2110 계열이 있습니다. SMPTE 2022는 10Gbps 네트워크에서 사용할 수 있는 비디오 데이터 전송방식입니다. 하지만 10Gbps의 전송 대역폭으로는 12Gbps의 UHD 영상 데이터를 전송할 수 없는 한계를 가지고 있습니다. 이 한계를 극복하기 위한 표준이 25Gbps를 지원하는 SMPTE 2110 표준입니다. SMPTE 2022와 SMPTE 2110을 들여다보기 위해서는 현재 방송제작에서 사용하는 비디오 데이터 전송방식인 SDI(Serial Data Interface)의 대략적인 구조에 대해서 알아야 합니다. 그래서 이번 연재에서는 SMPTE 2022와 SMPTE 2110 표준에 앞서 간략하게 SDI 신호 포맷에 대해 다루도록 하겠습니다. 본 연재를 통한 SDI 신호 포맷의 설명은 HD-SDI 포맷 중 1080i 방식 포맷의 설명을 시작으로 3G-SDI, 12G-SDI 등으로 확대해 나가겠습니다.

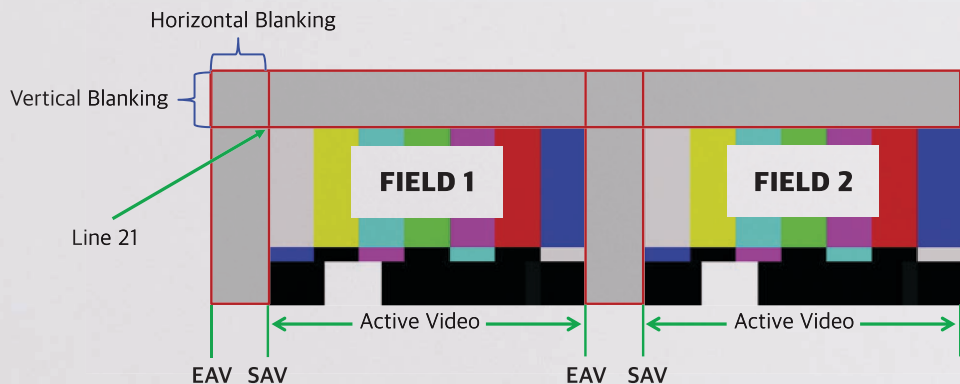


그림 1. HD-SDI 신호구조

[그림 1]은 1080i 포맷의 HD-SDI 신호를 나타낸 것입니다. 1080i 포맷은 Interlace 방식이므로 [그림 1]과 같이 영상의 홀수 번째 줄 데이터로 이루어진 FIELD 1과 짝수 번째 줄 데이터로 이루어진 FIELD 2가 번갈아 반복되는 구조입니다. [그림 1]의 Vertical Blanking은 과거 브라운관 TV에서 전자총(Cathode Ray Tube)으로부터 각 FIELD의 주사가 완료되었을 때 화면의 최하단을 향해 주사되는 전자들의 방향을 다시 화면의 최상단으로 옮기는데 필요한 시간에 대응되고, Horizontal Blanking은 영상의 한 라인에 대한 주사가 완료되었을 때 화면의 우측 끝으로 주사되는 전자들의 주사방향을 다음 라인의 주사를 위하여 좌측 끝으로 옮기는데 필요한 시간에 대응됩니다. 우리가 눈으로 보는 HD 영상은 1080개의 라인으로 이루어져 있지

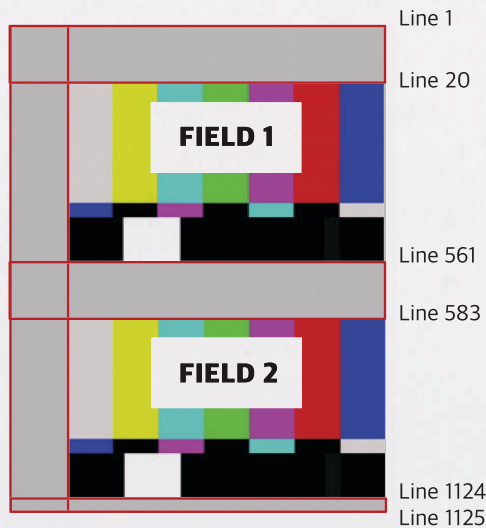


그림 2. Interlace 방식과 Vertical Blanking

만, Vertical Blanking이 45라인을 차지하고 있으므로 실제 영상 신호 속의 라인은 1125라인입니다. 그런데 예리하신 독자들은 뭔가 이상한 눈치를 채셨을 것 같습니다. [그림 1]에는 FIELD 1과 FIELD 2의 Vertical Blanking의 라인수가 같아 보이므로 45라인이 각 FIELD에 분배가 되어야 하는데 짝수가 아니므로 정확히 둘로 나뉘지 않습니다. 그리고 FIELD 1의 Active Video가 21번째 라인에서 시작하는 걸로 그려져 있어서 그림과 설명이 서로 맞지 않는 것으로 보입니다. 만약 Progressive 방식의 HD 포맷이라면 45라인의 Vertical Blanking 이후에 1080라인의 Active Video를 포함한 라인이 이어지므로 쉽게 1125라인의 구성이 가능하지만 Interlace 방식의 경우 정확하게는 [그림 2]와 같이 540라인의 FIELD 1을 시작하기 전에 Line 1~20에 해당하는 20개 라인의 Vertical Blanking 구간을 갖고, FIELD 1이 완료된 후에 Line 561~583에 해당하는 23개 라인의 Vertical Blanking 구간이 이어지며 또다시 540라인의 FIELD 2가 완료된 후에 Line 1124와 Line 1125 두 개의 추가적인 Vertical Blanking 구간이 붙습니다. 이렇게 총 45개 라인의 Vertical Blanking 구간과 540개 라인의 두 FIELD가 합쳐져서 1125라인을 형성하게 됩니다.

또한 HD-SDI의 Horizontal Blanking은 280화소에 해당하므로 수평 방향의 총 화소수도 1920개가 아닌 2200개입니다. 그러므로 1920×1080 사이즈의 Active Video를 전송하기 위해 SDI에서 쓰는 영상 데이터의 사이즈는 2200×1125가 됩니다. Vertical/Horizontal Blanking에 대해서 알게 되었으니, [그림 1]에 표시된 것 중에 설명이 필요한 건 EAV와 SAV만 남았습니다. EAV는 End of Active Video의 약자이며, SAV는 Start of Active Video의 약자입니다. EAV는 말 그대로 Active Video가 끝났음을 의미하므로 SAV까지는 비디오 관련 데이터가 없습니다. EAV와 SAV가 중요한 이유는 이들을 통해 FIELD와 Vertical/Horizontal Blanking 구간을 구별하기 때문입니다. EAV와 SAV를 구체적으로 설명하기 전에 자연스럽게 떠오르는 한 가지 의문을 짚고 넘어가겠습니다. 그 자연스러운 의문은 EAV와 SAV 사이는 아무것도 없는 공백인지 아닌지에 관한 것입니다. 사실 상식적으로 신호에 아무것도 없는 빈 구간을 두는 것이 왠지 이상하지요? 네, SDI 신호에서도 EAV와 SAV 사이를 그냥 아무 쓸모 없는 공백으로 남겨두지 않았습니다. Embedded Audio와 같은 부가데이터가 EAV와 SAV 사이에 들어갑니다. 이제 딱 보기만 해도 자연스럽게 떠오르는 의문에 대한 답을 얻었으니, 곰곰이 생각할수록 하나 둘 피어나는 의문들에 대한 답을 찾아가 보도록 하겠습니다. 그럼 [그림 1]을 약 5분 정도 보면서 다 같이 곰곰이 생각해보며 어떤 의문이 떠오르는지 살펴보는 시간을 갖도록 하겠습니다.

대략 5분이 경과했다고 가정하고, 5분 동안 [그림 1]을 보면서 아무거나 곰곰이 생각해보셨나요? 어떤 의문이 떠올랐나요? 모르긴 모르지만 아마도 독자 여러분의 생각 속에서 Active Video Data는 도대체 어떻게 생긴 것이고, EAV와 SAV를 어떻게 구분할까 하는 의문이 떠올랐을 것 같습니다. 제 예상이 거의 99% 맞았다고 치고 [그림 1]에 나타난 Active Video Data, EAV, SAV가 실제로 어떤 것인지 이야기해보겠습니다. HD 방송제작에 쓰이는 영상신호는 YCbCr 4:2:2 형태의 영상 데이터입니다. 카메라의 감광소자로부터 얻는 빛의 3원색에 대응되는 R, G, B 신호가 어떻게 YCbCr로 바뀌고, 이 YCbCr의 배열 중에 4:2:2는 어떤 배열인지에 관해서는 독자 여러분 모두 잘 알고 계시다고 간주하여 생략하겠습니다. 카메라에서 화소마

다 R, G, B가 YCbCr로 바뀌어 SDI 신호로 만들어지는 것을 개념적으로 표현하면 [그림 3]과 같습니다.

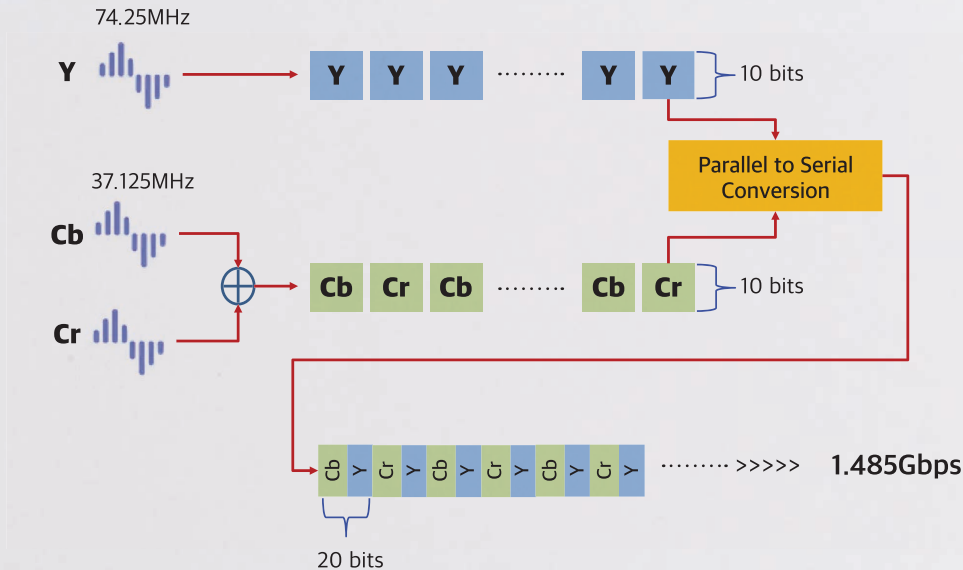


그림 3

[그림 3]에서 1080i 포맷 HD-SDI의 Y신호는 74.25MHz의 주파수로 샘플링됩니다. 74.25MHz의 샘플링 주파수가 사용되는 것은 간단한 산수를 통해서 계산이 가능합니다. 앞서 설명한 대로 1920×1080 해상도의 영상이라도 해도 Vertical/Horizontal Blanking 구간을 포함하면 실제 영상 데이터의 사이즈는 2200×1125이며, 60i는 초당 30프레임과 같으므로 1초 분량의 영상 데이터를 채우기 위해 필요한 Y샘플의 개수는 $2200 \times 1125 \times 30 = 74,250,000$ 개가 됩니다. 이는 초당 74.25MHz의 주파수로 Y값을 샘플링하는 것과 같은 이야기가 되므로 Y값의 샘플링 주파수는 74.25MHz가 되는 것이며, Cb/Cr값들은 화소를 하나씩 건너뛰며 샘플링을 하므로 Y값을 샘플링하는 주파수인 74.25MHz의 절반인 37.125MHz의 주파수로 Cb/Cr 각각을 샘플링하게 됩니다. 그런데 여기서 SDI(Serial Data Interface)라는 이름이 말하듯이 최종 출력을 Y값 따로, Cb/Cr값 따로 병렬로 출력할 수는 없습니다. Y값과 Cb/Cr값을 시리얼(Serial)로 묶어서 하나의 비트 스트림으로 출력해야 합니다. 그러므로 각각 10비트로 샘플링된 Y, Cb, Cr값을 시리얼 데이터로 만들어서 출력하면 $2(Y \text{와 } Cb/Cr) \times 10(\text{bits}) \times 74,250,000 = 1,485,000,000\text{bps} = 1.485\text{Gbps}$ 가 됩니다(bps는 bits per second의 약자). [그림 3]을 통해 방송 제작에 쓰이는 SDI 영상 신호가 YcbCr 4:2:2 포맷이며, Y, Cb, Cr은 각각 10비트이고, 신호의 대역폭은 1.485Gbps인 것이 서로 어떻게 얹혀있는지 쉽게 이해하실 수 있었을 것 같습니다.

그렇다면 이제 남은 궁금증인 EAV, SAV에 관해서 알아보도록 하겠습니다. EAV와 SAV는 4개 화소에 걸친 Y와 Cb/Cr 값의 자리에 3FF 000 000 XYZ를 채우는 것으로 정의되어 있습니다. 혹시 3FF나 000 같은 표현이 어색한 독자분이 계시지 모르므로 3FF 같은 표기법에 관해 잠시 설명하겠습니다. 3FF와 같은 표시 방식은 이진코드를 4비트씩 끊어서 읽는 방법입니다. 4비트의 이진 코드로 표현할 수 있는 수의 가지 수는 $2^4=16$ 개이므로 0~15까지의 숫자를 표현할 수 있습니다. 그래서 이진 코드를 4비트씩 끊어 읽을 경우 0부터 9까지는 한 자리 숫자 그대로 쓰고 10~15까지를 A~F의 알파벳을 대응시켜서 표시하게 됩니다. 예를 들면, 3FF를 이진 코드로 쓰면 3은 11, F는 1111이 되어 3FF = 1111111111이 되며

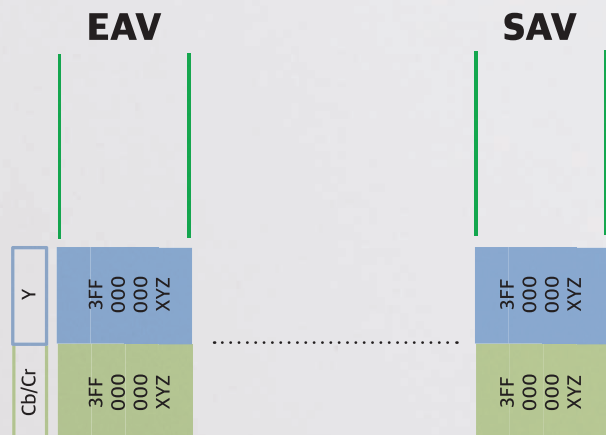


그림 4. EAV와 SAV 비트스트림

십진수로 1023이 됩니다. 000은 10비트의 이진 코드로 당연히 0000000000와 같이 10개의 0이 됩니다. SDI 신호에서 Active Video 구간의 화소에서는 Y나 Cb/Cr값은 3FF와 000 값을 가질 수 없게 설계되어 있습니다. 그러므로 3FF 000 000가 연속하여 나타나면 EAV 또는 SAV가 되는 것을 쉽게 검출할 수 있습니다.

그렇다면 3FF 000 000 다음에 나오는 XYZ는 무엇일까요? XYZ는 위에서 설명한 4비트로 끊어 읽기와도 상관없는 문자들이니 이진 숫자를 표현한 것 같지도 않습니다. XYZ는 [표 1]에 정리된 바와 같이 10비트의 코드에 붙인 이름이며 각 비트를 통해 특정 정보를 전달하도록 설계되어 있습니다. [표 1]에 정리된 바와 같이 XYZ에서 F로 표시된 비트는 현재 SDI 신호가 FIELD 1에 해당하는지 FIELD 2에 해당하는지를 나타내며, V로 표시된 비트는 Vertical Blanking인지 아닌지를 나타내고, H로 표시된 비트는 3FF 000 000을 뒤따르는 XYZ가 EAV에 해당하는지 SAV에 해당하는지를 나타냅니다. P0~P3는 XYZ의 에러 검출을 위한 비트들입니다. [표 1]의 정리된 바와 같이 XYZ를 통해 쉽게 SDI 신호 내의 비디오 데이터를 분석할 수 있습니다.

비트	10 (MSB)	9	8	7	6	5	4	3	2	1 (LSB)
XYZ	1	F	V	H	P3	P2	P1	P0	0	0

F : 0 = FIELD 1, 1 = FIELD 2

V : 0 = Vertical Blanking 아님, 1 = Vertical Blanking

H : 0 = SAV, 1 = EAV

MSB : Most Significant Bit

LSB : Least Significant Bit

P0~P3 : XYZ 에러 검출 관련 비트

표 1. XYZ 비트별 정보

TMI(Too Much Information)를 덧붙이면 Y 및 Cb/Cr 신호에 삽입된 3FF 000 000 XYZ가 시리얼 데이터로 최종 출력되면 [그림 5]와 같은 형태로 HD-SDI 신호를 통해 출력됩니다.

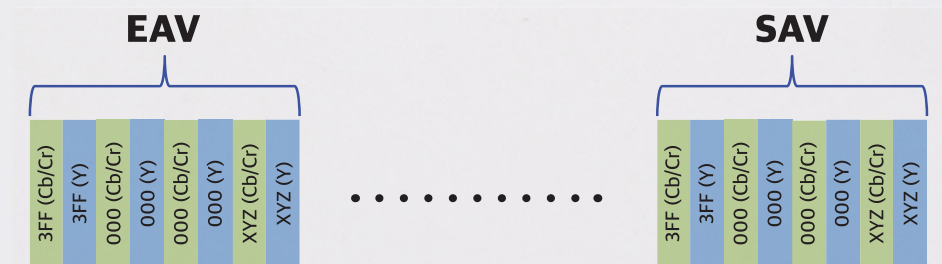


그림 5. 시리얼 데이터로 전달되는 EAV와 SAV

지금까지 HD-SDI 신호 포맷에 관해 간략히 다루어 보았습니다. 다음 연재에서는 UHD를 위한 3G-SDI 및 12G-SDI 등에 관해 이야기해보겠습니다. 추워지는 날씨에 독자 여러분 모두 건강 유의하시기 바랍니다. 🌧️