

## Multi-Screen Service Forum Specification

MSS.S-Y16-001

제정일: 2016년 08월 10일

HMD를 이용한 VR360 화상회의 시스템에서의  
시선정보 시그널링  
User Sight Information Signaling for  
VR360 Video Conference System using HMD

제출일 : 2016년 7월 31일  
제출기관 : 멀티스크린서비스포럼  
제출인 : 류은석

## 서 문

## 1. 표준의 목적

이 표준의 목적은 최근 부상하고 있는 VR(Virtual Reality) 기술과 HMD(Head-Mounted Display)를 활용한 화상회의 시스템에서의 사용자 시선 정보를 시그널링(Signaling) 함으로써, 비디오 요구 대역폭을 낮추고, 사용자에게 빠르게 응답하는 기술을 설명함에 있다.

## 2. 주요 내용 요약

이 표준은 HMD를 통한 VR 관련 서비스를 제공할 때, 사용자의 시선 정보를 시그널링하는 기술을 담고 있으며, 기술의 장점을 설명하기 위해 (1) 본 표준의 필요성 및 스케일러블 비디오 및 HMD의 배경지식, (2) 본 표준을 HMD를 활용한 VR 화상회의 시스템에 적용했을 때 기대되는 성능향상, 그리고, (3) 제안하는 표준 시그널링 규격(신텍스, 시멘틱스)을 기술한다.

## 3. 인용 표준과의 비교

## 3.1 인용 표준과의 관련성

이 표준은 국제 표준단체 The Joint Collaborative Team on Video Coding (JCT-VC)의 Scalable Extension of High Efficiency Video Coding (SHVC) 기반 기술 등을 이용하는 시스템을 위한 별도의 독립적인 시그널링 표준으로써, SHVC 표준과 직접적인 관련성이 없음.

## Preface

### 1 Purpose

The purpose of this standard is to define the signaling syntax and semantics of the user sight information for VR360 video conference system using HMD. The signaling standard can decrease the video adaptation delay for user's movement as well as the bandwidth requirements of video transmission.

### 2 Summary

The standard includes the specifications for the user sight information signaling for VR service using HMD. It consists of three sections: (1) the motivation of this standard with the background of scalable video and HMD, and (3) the standard signaling specifications including syntax and semantics.

### 3 Relationship to Reference Standards

The standard can use the referenced video coding standard specifications such as the Scalable High Efficiency Video Coding (SHVC) of the Joint Collaborative Team on Video Coding (JCT-VC). But, this standard does not directly affect to or influenced by the referenced standard but specifies the signaling details independently.

## 목 차

1 적용 범위 .....	4
2 인용 표준 .....	4
3 용어 정의 .....	4
4 약어 .....	4
5 HMD를 이용한 VR360 화상회의 시스템에서의 시선정보 시그널링 .....	5
5.1 본 표준의 필요성 및 배경지식 .....	5
5.2 본 표준 기술을 통해 기대되는 성능향상 .....	6
5.3. 제안하는 표준 시그널링 규격 .....	7
5.4. 추가 확장적 사용 .....	9

# HMD를 이용한 VR360 화상회의 시스템에서의 시선정보 시그널링

## 1. 적용범위

본 표준의 적용 범위는 비디오 통신에서의 시그널링을 처리하는 객체를 다루며, 이는 사용자 단말, 서버, 중계 시스템 및 라우터 등을 포함한다. 또한, 본 표준의 시그널링 메시지는 (1) Session 정보를 실어나르는 high-level syntax 프로토콜을 통해 전해질 수도 있고, (2) 비디오 표준의 SEI, VUI, 또는 슬라이스 헤더 (Slice Header) 등의 패킷 단위에서 전해질 수도 있고, (3) 비디오 파일을 설명(Describe)하는 별도의 파일로(예: DASH의 MPD) 전달될 수 있다.

## 2. 인용 표준

"H.265: High efficiency video coding". ITU. 2015-07-09. Retrieved 2015-08-02.

## 3. 용어 정의

해당사항 없음

## 4. 약어

- VR: Virtual Reality
- HMD: Head-Mounted Display
- BL: Base Layer
- EL: Enhancement Layer
- HEVC: High Efficiency Video Coding
- UHD: Ultra High Definition
- MPEG: Moving Picture Experts Group
- JCTVC: The Joint Collaborative Team on Video Coding
- SEI: Supplemental Enhancement Information
- VUI: Video Usability Information
- DASH: Dynamic Adaptive Streaming over HTTP
- MPD:Media Presentation Descriptio
- SVC: Scalable Video Coding
- SHVC: Scalable HEVC
- ROI: Region of Interest
- CU: Coding Unit

## 5. HMD를 이용한 VR360 화상회의 시스템에서의 시선정보 시그널링

본 섹션에서는 HMD를 이용한 화상회의 시스템에서의 시선 정보 시그널링 표준에 관하여 3가지를 구체적으로 설명한다: (1)필요성, (2)기대되는 성능향상, (3)구체적인 시그널링 규격.

### 5.1. 본 표준의 필요성 및 배경지식

최근 VR 기술 및 장비의 발달과 함께 여러 기업을 통해 HMD와 같은 착용 가능한 기기들이 선보이고 있다. 이를 통한 여러 서비스 시나리오 중에는 가장 대표적인 영화 관람 및 게임뿐만 아니라 화상회의와 원격 수술 등이 있다.

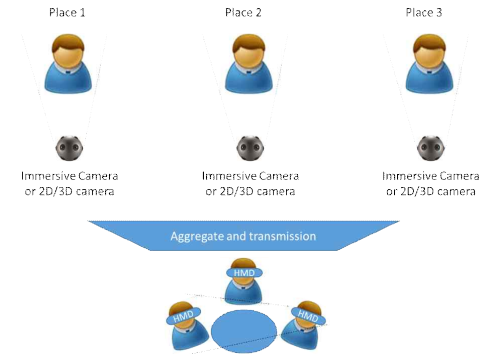


그림 5-1. HMD를 통한 VR 화상회의의 개념적 구성

이 중 화상회의의 서비스는 그림 5-1과 같이 원격 장소에 위치한 회의 참가자를 일반 카메라나 Immersive Camera 장비들을 이용하여 촬영하고, 이를 화상회의의 시스템에 전송하여 가상 공간 영역에 맵핑(Mapping)한다.

현재, 이 HMD를 이용한 화상회의의 시스템이 겪는 어려운 문제는 사용자 눈에는 매우 넓게 보이는 (1) 360 영상 전체를 담는 비디오 화소수가 매우 높아야 한다는 점이다. 따라서 UHD 급 영상을 이용할 필요성이 있는데, 이 경우 사용자 단말 및 화상회의의 중계 시스템의 (2)대역폭 확보가 어려운 문제점과, 처리해야 할 많은 비디오 데이터로 인해 (3)사용자의 머리 움직임에 빠르게 응답하기 어렵다는 문제점을 갖게 된다.

본 표준은 이의 해결책으로 '시선 정보 시그널링' 기법의 내용과 장점, 그리고 이를 통한 (1) 전체 대역폭을 낮춤, (2) 빠른 사용자 응답속도 지원을 설명한다.

#### 5.1.1. 배경지식

국제 비디오 코딩 표준 기술 중 스케일러블 비디오(Scalable Video)가 있다. SVC 및 HEVC의 확장 표준인 SHVC는 각각 기본 계층 (BL)와 하나 이상의 향상 계층 (EL)로 이루어져 있으며, 기본 계층만 받았을 때엔 일반 화질을 제공하고 향상 계층까지 함께 받을 경우 고 화질을 제공할 수 있다. 즉, 기본 계층과 하나 이상의 향상 계층이 있을 때, 기본 계층을 받은 상태에서 향상 계층을 더 받으면 받을수록 화질이나 제공하는 영상의 품질이 좋아지게 한다.

VR360 영상의 높은 대역폭을 낮추기 위해, 삼성전자는 화면의 상하단 영상에 대해 다운스케일링(down-scaling)을 하여 전체 비트레이트를 낮추고, 독일의 프라운호퍼 HHI에서는 SHVC기술을 사용하되, 기본/향상 계층 모두에서 타일 기법을 사용하여 필요한 부분만 전송하고 업데이트 하는 기술을 사용한다. 하지만, 특히 후자의 기술은 빠른 사용자 응답속도를 지원하기 어렵다는 문제점을 가진다.

## 5.2. 본 표준 기술을 통해 기대되는 성능향상

### 5.2.1. 대역폭 문제의 해결

전체 영상을 하나의 압축된 영상 비트스트림(Bitstream)으로 받아서 이를 디코딩(decoding)하고 사용자가 바라보는 영역을 가상의 공간에 렌더(render)하는 기술은 현재 대부분 전체 영상(예: 360도 immersive 영상)을 모두 비트스트림으로 전송 받는다. 각각이 고해상도인 영상이 모인 이 비디오 비트스트림의 총 대역폭은 매우 클 수밖에 없고 이를 막기 위해서 국제 비디오 표준 기술 중 SVC 및 HEVC의 스케일러블 확장 표준인 SHVC와 같은 스케일러블 비디오 기술이 사용될 수 있다. 그 방법은 그림 5-2와 같다.

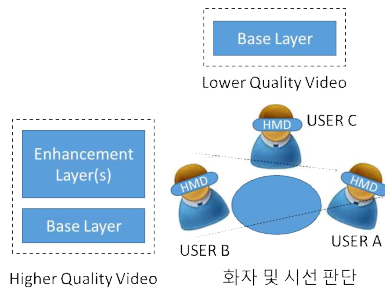


그림 5-2. 스케일러블 비디오 기술을 활용한 VR 화상회의

위의 그림에서 사용자(User) A가 다른 사용자 B와 C와 함께 가상 공간에서 화상회의를 할 때, (1)사용자가 가상공간에서 HMD를 이용하여 바라보는 시점 및 줌(Zoom) 영역 등의 영상 구성을 파악한다. 기술적으로는 HMD 또는 외부의 센서 등을 통해 확인한다. (2)예를 들어 User A만 바라보거나 User C가 화면에서 아주 작은 영역만 차지하고 있을 경우, User B의 정보는 스케일러블 비디오 코딩에서 기본 계층과 향상 계층 모두를 보내주고, User C는 기본 계층 정보만 보내줌으로써 전체 대역폭을 낮출 수 있다.

### 5.2.2. 사용자 시선 움직임에 실시간 반응하는 비디오 프로세싱 속도 문제의 해결

기존 방법은 User A가 User B를 바라보다가 User C로 고개를 돌렸을 때, HMD의 센서 등으로 이 움직임을 파악하여 재빨리 User C를 표현하기 위한 비디오 정보를 처리하고 화면에 재생한다. 문제는 매우 빨리 새로운 영역의 영상을 처리하는 것이 어렵기에 지금까지 모든 데이터를 다 미리 받아두는 비효율적 방법을 대부분 취해왔다.

본 표준 기술은 그림 5-3과 같이 User A가 User C로 고개를 돌렸을 때 가지고 있는 기본 계층 데이터를 이용하여 빠르게 사용자에게 응답한다. 전체 고화질 데이터를 처리할 때 보다 더 빨리 재생할 수 있기에 빠른 응답에 좋다. 다만, 기본 계층의 화질은 향상 계층까지

받았을 때 보다 낮으므로, 그림 5-3(우)와 같이 사용자가 고개를 돌린 정보를 센서 등으로부터 얻는 순간 바로 그 부분의 영상에 대한 향상 계층 (주로 SHVC의 타일 기법을 활용한)을 전송하여 짧은 시간 내에 고화질 영상으로 서비스한다.

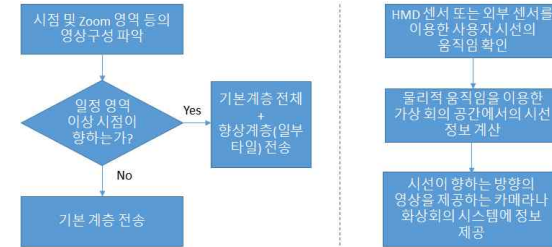


그림 5-3. 사용자 시선 움직임에 따른 전송할 영상 계층의 선택 (좌: 본 표준의 계층선택 판단 기준, 우: 시선 정보 판단 및 내용 전달 방법)

이 방법은 사전에 일부 추가 영역의 데이터만 전송 받아두는 단순 프리캐칭(precaching) 기법이나 아예 시선이 바라보는 영역만의 데이터를 전송 받는 방법에 비해 큰 장점을 가진다.

## 5.3. 제안하는 표준 시그널링 규칙

### 5.3.1. 선택적 시선 시그널링

앞 절에서 설명된 이유로 사용자의 시선 정보를 시그널링 하는 부분은 매우 중요하다. 하지만, 이 정보가 센싱되고 단말간에 전달되는 주기가 너무 잦을 경우 전체 네트워크 및 사용자 단말 또는 화상회의 시스템에 부담을 준다. 따라서, 본 발명은 그림 5-4와 같이 사용자의 시선이 '미땅히 주목 받을 것으로 예상되는 (본 표준에서는 ROI로 표기함)' 오브젝트를 향할 경우 시그널링 하지 않고, 다른 곳을 향할 때에만 시그널링 하는 선택적 시그널링을 사용한다. 이 때 ROI에 해당한다고 판단되는 오브젝트는, 화상회의에서 현재 시점의 화자(말하는 이), 비록 말을 하고 있는 화자는 아니지만 퍼포먼스를 하고 있거나 (온라인 강의의 경우) 칠판에 무엇인가를 쓰는 이 (참고로, 시스템상의 옵션(예: 강의자는 User B)을 통해 정보를 받을 수도 있다), 또한, VR 게임에서는 새로 등장하는 상대 오브젝트 동일 수 있다. 이 때, 시선이라 함은 다음 그림과 같이 실제 공간이 아닌 가상 공간에서의 시선이며, 따라서 사용자를 중심으로 가상의 화상회의 공간에 위치한 다른 이를 바라보는 동작을 센싱하여 처리한다.

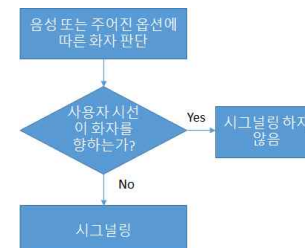


그림 5-4. 본 표준의 선택적 시선 정보 시그널링

5.3.2. 스케일러블 비디오에서의 ROI 시그널링

기본 계층은 빠른 사용자 응답시간을 위해 타일링되지 않고 전체적으로 인코딩된다. 하나 이상의 향상계층은 필요에 따라 일부 또는 전체가 여러 타일들로 나누어져서 인코딩된다. 이 때, ROI는 (1) 사용자가 보게될 가상의 공간에서 중요 오브젝트가 위치할 타일 위치 (예: 게임 등에서 새로운 적이 등장하는 위치, 화상 통신에서 가상공간에서의 화자의 위치), 또는 (2) 사용자의 시선이 바라보는 곳에 해당할 수 있다.

ROI의 타일 번호는 연속적이므로 모든 번호를 다 보내지 않고도 효과적으로 압축할 수 있다 (예: 타일 시작과 끝 번호, 좌표점 정보, 타일 내 CU번호 리스트, 타일 번호를 수식으로 표현 등).

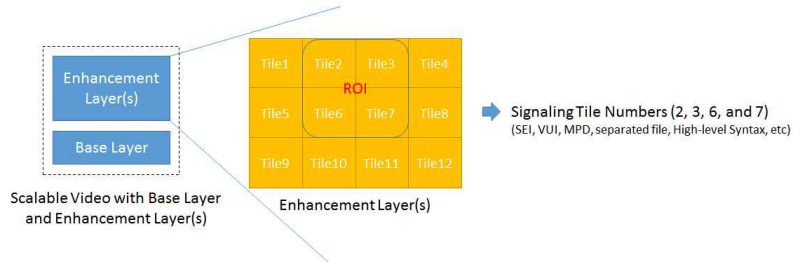


그림 5-5. 본 표준의 ROI 정보 시그널링 예

이 시그널링은 그림 5-5와 같이 (1) Session 정보를 실어나르는 하이레벨 신택스 (High-level syntax) 프로토콜을 통해 전해질 수도 있고, (2) 비디오 표준의 SEI, VUI, 또는 슬라이스 헤더 등의 패킷 단위에서 전해질 수도 있고, (3) 비디오 파일을 설명하는 별도의 파일로(예: DASH의 MPD) 전달될 수 있다.

위의 시그널링을 통해 향상계층의 필요 타일만 기기간 또는 서버와 클라이언트간 전송/전달 받음으로써 전체적인 대역폭을 낮추고, 비디오 프로세싱 시간을 줄일 수 있다. 이는 빠른 HMD 사용자 응답시간을 보장하는데 매우 중요하다.

표 5-1과 표 5-2는 본 표준에서 제안하는 비디오 픽처별 시선정보(ROI) 시그널링 규격과 파일, 청크, 비디오 픽처 그룹별 시선정보(ROI) 시그널링 규격을 다루고 있다.

표 5-1. 제안하는 비디오 픽처별 시그널링 규격

Syntax	Semantics
정보 모드 (1-Byte)	담고 있는 정보 모드 (1번 - 5번) 0: 이전 모드 정보 그대로 사용 (제외될 수 있음) 1: 모든 타일 정보 리스트 2: 연속된 타일의 시작과 끝 번호 3: 좌상단 및 우하단 타일 번호 4: 타일 번호 및 타일에 포함된 CU 번호 리스트 5: 사용자 정의 모드 사용 (Reserved)
데이터 크기 (7-Byte)	본 시그널링 정보를 담는 데이터 크기
타일 번호 리스트	1-3번 모드

(타일 번호, CU번호 리스트) 리스트	4번 모드
사용자 정의 모드 (정의된 데이터 크기 기준)	사용자가 추가/선택적으로 전송하려는 정보(들)

표 5-2. 제안하는 파일, 청크, 비디오 픽처 그룹별 시그널링 규격

Syntax	Semantics	
버전 정보	시그널링 규격 버전 정보	
전체 데이터 크기	본 시그널링 데이터 전체크기 (패킷 및 파일일 수 있음)	
단위 정보	픽처 번호	비디오 픽처 번호 예: HEVC에서는 POC(Picture Order Count)가 될 수 있으며, 일반 비디오 코덱의 경우 해당 픽처(프레임) 번호 담고 있는 정보 모드 (1번 - 5번) 0: 이전 모드 정보 그대로 사용 (제외될 수 있음) 1: 모든 타일 정보 리스트 2: 연속된 타일의 시작과 끝 번호 3: 좌상단 및 우하단 타일 번호 4: 타일 번호 및 타일에 포함된 CU 번호 리스트 5: 사용자 정의 모드 사용 (Reserved)
	정보 모드 (1-Byte)	본 시그널링 정보를 담는 데이터 크기
	데이터 크기 (7-Byte)	본 시그널링 정보를 담는 데이터 크기
	타일 번호 리스트	1-3번 모드
	(타일 번호, CU번호 리스트) 리스트	4번 모드
	사용자 정의 모드 (정의된 데이터 크기 기준)	사용자가 추가/선택적으로 전송하려는 정보(들)
새로운 단위 정보 (반복)	단위 정보 반복: 여러 픽처별 타일 정보가 나온 단위가 픽처 묶음에 대해 반복될 수 있음.	

5.4. 추가 확장적 사용

본 표준은 스케일러블 비디오를 통한 차별적 전송 기법을 이야기하고 있지만, 단일 계층을 갖는 일반 비디오 코딩 기술을 사용할 경우에도 다음 그림과 같이 양자화 계수 (Quantization Parameter)나 다운/업 스케일링 정도를 조절함으로써 (1) 전체 대역폭을 낮추고, (2) 빠르게 사용자 시선 움직임에 응답하는 등의 장점을 보일 수 있으며, MPEG DASH, MS의 Smooth Streaming, Apple의 HLS와 같이 미리 여러 비트레이트를 갖는 비트스트림으로 트랜스코딩 된 파일들을 사용할 경우 청크(Chunk) 단위로 높은 수준과 낮은 수준 사이를 스위치 할 수 있다.

## 표준작성 공헌자

포럼 표준 번호 : MSS.S-Y16-001

이 표준의 제개정 및 발간을 위해 아래와 같이 여러분들이 공헌하였습니다.

구분	성명	위원회 및 직위	연락처	소속사
표준(과제)제안	류은석		031-750-8905 010-4893-2199	가천대학교
표준 초안 작성자	류은석		031-750-8905 010-4893-2199	가천대학교
	류영일		031-750-8905	가천대학교
표준안 심의				
사무국 담당				