

TTA Standard

(기술보고서는 TTA Technical Report)

정보통신단체표준(국문표준)

TTAx.xx-xx.xxxx/R1

제정일: 200x 년 xx 월 xx 일

개정일: 200x 년 xx 월 xx 일

머리장착형영상장치를 이용한 가상
현실 화상회의 시스템에서의 시선정
보 시그널링

User Sight Information Signaling for Virtual
Reality Video Conference System using
Head-Mounted Display (HMD)



한국정보통신기술협회
Telecommunications Technology Association

표준초안 검토 위원회	디지털콘텐츠 프로젝트그룹(PG610)				
표준안 심의 위원회	소프트웨어/콘텐츠 기술위원회(TC6)				
	성명	소속	직위	위원회 및 직위	표준번호
표준(과제) 제안	류은석	가천대학교	조교수	-	
표준 초안 작성자	류은석	가천대학교	조교수	-	
	류영일	가천대학교	대학원생	-	
	노현준	가천대학교	대학원생	-	
사무국 담당				-	

본 문서에 대한 저작권은 TTA에 있으며, TTA와 사전 협의 없이 이 문서의 전체 또는 일부를 상업적 목적으로 복제 또는 배포해서는 안 됩니다.

본 표준 발간 이전에 접수된 지식재산권 확약서 정보는 본 표준의 '부록(지식재산권 확약서 정보)'에 명시하고 있으며, 이후 접수된 지식재산권 확약서는 TTA 웹사이트에서 확인할 수 있습니다.

본 표준과 관련하여 접수된 확약서 외의 지식재산권이 존재할 수 있습니다.

발행인 : 한국정보통신기술협회 회장

발행처 : 한국정보통신기술협회

13591, 경기도 성남시 분당구 분당로 47

Tel : 031-724-0114, Fax : 031-724-0109

발행일 : 20xx.xx

서 문

1 표준의 목적

이 표준의 목적은 최근 부상하고 있는 가상 현실(Virtual Reality) 기술과 머리장착형영상장치(Head-Mounted Display; HMD)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 머리장착형영상장치를 통한 가상 현실 관련 서비스를 제공할 때, 사용자의 시선 정보를 시그널링하는 기술을 담고 있으며, 기술의 장점을 설명하기 위해 (1) 본 표준의 필요성 및 스케일러블 비디오 및 머리장착형영상장치의 배경지식, (2) 본 표준을 머리장착형영상장치를 활용한 가상 현실 화상회의 시스템에 적용했을 때 기대되는 성능향상, 그리고, (3) 제안하는 표준 시그널링 규격(신텍스, 시멘텍스)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	1
2	인용 표준	1
3	용어 정의	1
4	약어	1
5	머리장착형영상장치를 이용한 가상현실 화상회의 시스템에서의 시선정보 시그널링	2
5.1	본 표준의 필요성	2
5.2	본 표준 기술을 통해 기대되는 성능향상	3
5.3	제안하는 표준 시그널링 규격	4
5.4	추가 확장적 사용	9
부록	1-1 지식재산권 요약서 정보	10
	1-2 시험인증 관련 사항	11
	1-3 본 표준의 연계(family) 표준	12
	1-4 참고 문헌	13
	1-5 영문표준 해설서	14
	1-6 표준의 이력	15

머리장착형영상장치를 이용한 가상현실 화상회의 시스템에서의 시선정보 시그널링

1 적용 범위

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) 세션(Session) 정보를 실어나르는 고수준 구문(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

해당 사항 없음

4 약어

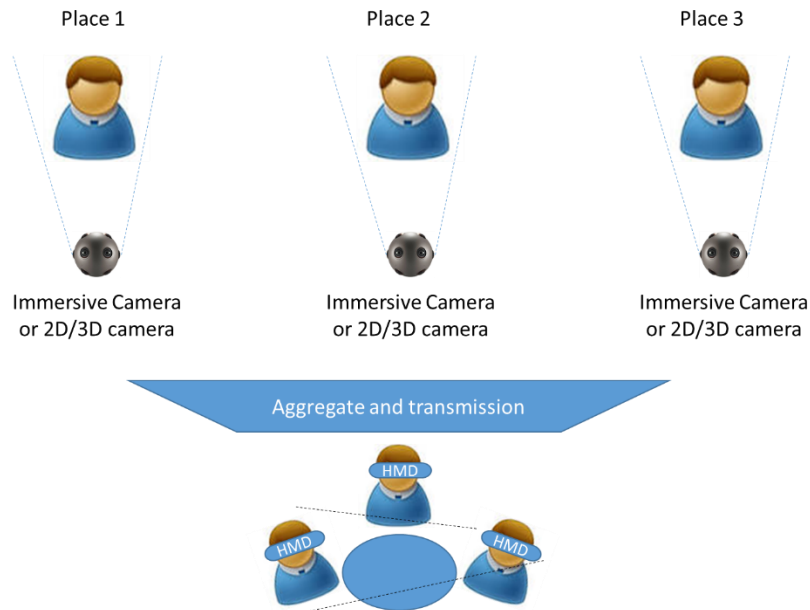
- BL: Base Layer
- EL: Enhancement Layer
- HEVC: High Efficiency Video Coding
- UHD: Ultra High Definition
- MPEG: Moving Picture Experts Group
- JCTVC: The Joint Collaborative Team on Video Coding
- SEI: Supplemental Enhancement Information
- VUI: Video Usability Information
- DASH: Dynamic Adaptive Streaming over HTTP
- MPD: Media Presentation Description
- ROI: Region of Interest
- CU: Coding Unit

5 머리장착형영상장치를 이용한 가상현실 화상회의 시스템에서의 시선정보 시그널링

본 섹션에서는 머리장착형영상장치(HMD)를 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

최근 가상 현실 기술 및 장비의 발달과 함께 여러 기업을 통해 머리장착형영상장치와 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 화상회의 서비스는 그림 5-1과 같이 원격 장소에 위치한 회의 참가자를 일반 카메라나 몰입형 카메라 (Immersive Camera) 장비들을 이용하여 촬영하고, 이를 화상회의 시스템에 전송하여 가상 공간 영역에 맵핑(Mapping)한다.

현재, 이 HMD를 이용한 화상회의 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 화상회의 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

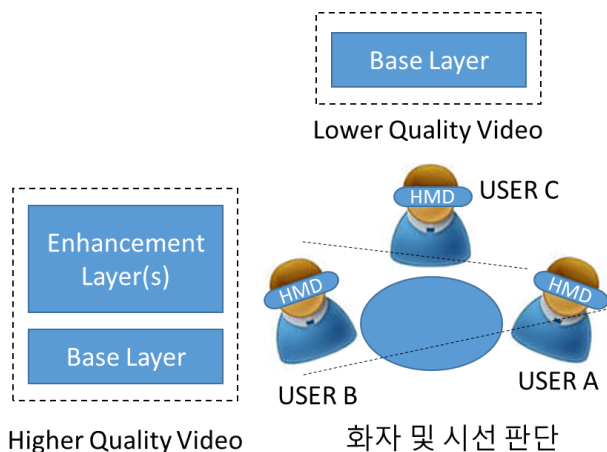
국제 비디오 코딩 표준 기술 중 스케일러블 비디오(Scalable Video)가 있다. 스케일러블 비디오 부호화 (Scalable Video Coding) 및 고효율 비디오 부호화 (High Efficiency Video Coding)의 확장 표준인 스케일러블 고효율 비디오 부호화 (Scalable High Efficiency Video Coding)는 각각 기본 계층(BL)과 하나 이상의 향상 계층 (EL)으로 이루어져 있으며, 기본 계층만 받았을 때엔 일반 화질을 제공하고 향상 계층까지 함께 받을 경우 고품질을 제공할 수 있다. 즉, 기본 계층과 하나 이상의 향상 계층이 있을 때, 기본 계층을 받은 상태에서 향상 계층을 더 받으면 받을수록 화질이나 제공하는 영상의 품질이 좋아지게 한다.

가상현실 회의 영상의 높은 대역폭을 낮추기 위해, 삼성전자는 화면의 상하단 영상에 대해 다운스케일링(down-scaling)을 하여 전체 비트레이트(Bitrate)를 낮추고, 독일의 프라운호퍼 HHI에서는 스케일러블 고효율 비디오 부호화 (Scalable High Efficiency Video Coding)기술을 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 복호화(decoding)하고 사용자가 바라보는 영역을 가상의 공간에 렌더(render)하는 기술은 현재 대부분 전체 영상(예: 360도 몰입형(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 스케일러블 고효율 비디오 부호화 (Scalable High Efficiency Video Coding)와 같은 스케일러블 비디오 기술이 사용될 수 있다. 그 방법은 그림 5-2와 같다.



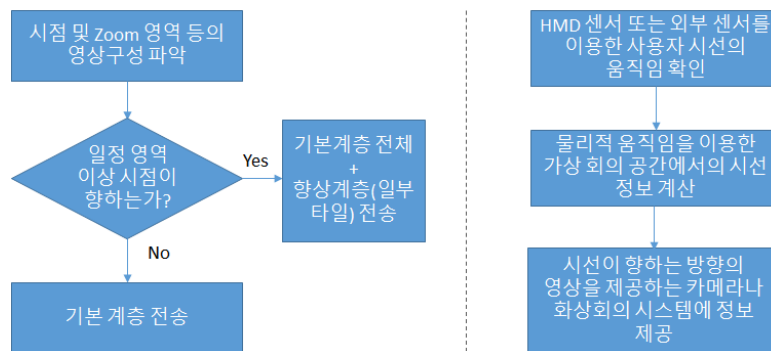
(그림 5-2) 스케일러블 비디오 기술을 활용한 VR 화상회의

위의 그림에서 사용자(User) A가 다른 사용자 B와 C와 함께 가상 공간에서 화상회의를 할 때, (1)사용자가 가상공간에서 머리장착형영상장치를 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 보단 낮으므로, 그림 5-3(우)와 같이 사용자가 고개를 돌린 정보를 센서 등으로부터 얻는 순간 바로 그 부분의 영상에 대한 향상 계층 (주로 스케일러블 고효율 비디오 부호화 (Scalable High Efficiency Video Coding)의 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)

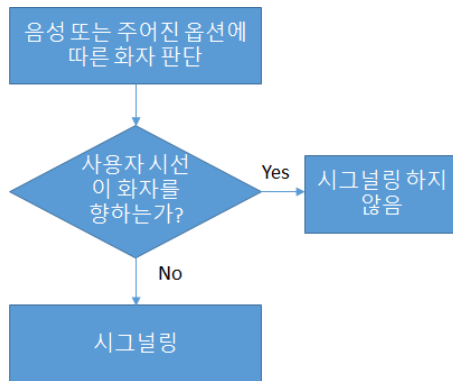
이 방법은 사전에 일부 추가 영역의 데이터만 전송 받아두는 단순 pre-caching 기법이 나 아예 시선이 바라보는 영역만의 데이터를 전송 받는 방법에 비해 큰 장점을 가진다.

5.3. 제안하는 표준 시그널링 규격

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 네

트위크 및 사용자 단말 또는 화상회의 시스템에 부담을 준다. 따라서, 본 발명은 그림 5-4와 같이 사용자의 시선이 ‘마땅히 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 등일 수 있다. 이 때, 시선이라 함은 다음 그림과 같이 실제공간이 아닌 가상공간에서의 시선이며, 사용자를 중심으로 가상의 화상회의 공간에 위치한 다른 이를 바라보는 동작을 센싱하여 처리한다.

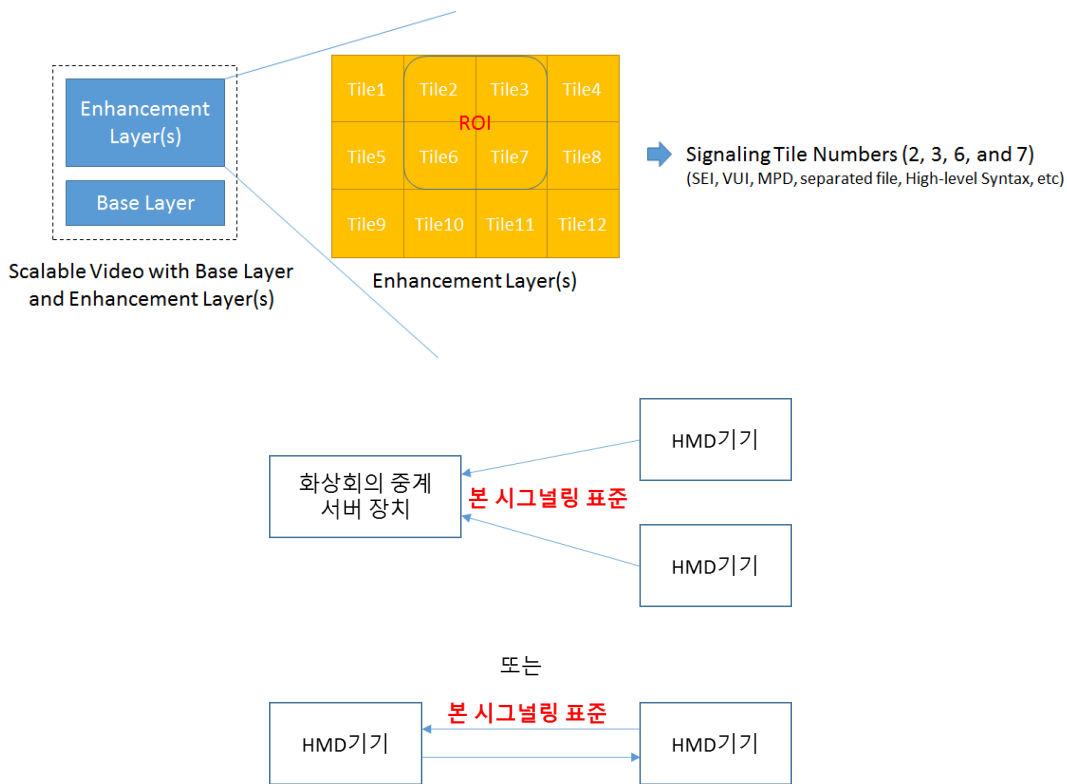


(그림 5-4) 본 표준의 선택적 시선 정보 시그널링

5.3.2. 스케일러블 비디오에서의 관심 영역 (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, 관심 영역(ROI)는 (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.

관심 영역(ROI)의 타일 번호는 연속적이므로 모든 번호를 다 보내지 않고도 효과적으로 압축할 수 있다 (예: 타일 시작과 끝 번호, 좌표점 정보, 타일 내 코딩 단위(CU) 번호 리스트, 타일 번호를 수식으로 표현 등).



(그림 5-5) 본 표준의 관심영역 정보 시그널링 예

이 시그널링은 그림 5-5와 같이 (1) 세션(Session) 정보를 실어나르는 고수준 구문 (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 5-1은 H.264 AVC나 H.265 HEVC와 같은 국제 비디오 표준에서의 SEI 메시지 페이로드 (payload) 구문 (Syntax)의 예 (sighted_tile_info)를 보여주고 있다. 만일 제안하는 구문이 185번으로 정해진 경우 다음과 같다. 검은색 글씨는 기존의 표준 구문이며, 붉은 색 글씨는 새로 추가되어야 할 내용이다.

<표 5-1> 제안하는 SEI 페이로드 구문의 예

Sei_payload(payloadType, payloadSize) {	Descriptor
if(payloadType == 0)	
buffering_period(payloadSize)	
.....	

else if(payloadType == 185)	
sighted_tile_info(payloadSize)	
.....	

다음 표 5-2와 표 5-3은 본 표준에서 제안하는 비디오 픽처별 관심 영역(시선 정보)(ROI) 시그널링 규격과 파일, 청크, 비디오 픽처 그룹별 관심 영역(시선 정보)(ROI) 시그널링 규격을 다루고 있고, 표 5-4는 이의 구문 설명을 한다.

표에 나온 u(n)는 통상 프로그래밍 언어에서 부호가 없는 (unsigned) ‘n’ 비트 수를 의미하며, ‘v’로 표시된 부분은 변화 가능한 비트수(표준에서는 varies로 읽힘)를 의미한다.

<표 5-2> 제안하는 비디오 픽처별 시그널링 규격

sighted_tile_info(payloadSize) {	비트수
info_mode	u(4)
tile_id_list_size	u(8)
for (i=0; i < list_id_list_size; i++) {	
tile_id_list [i]	u(8)
}	
if (info_mode == 4) {	
cu_id_list_size	u(16)
cu_id_list [i]	u(16)
}	
user_info_flag	u(1)
if (user_info_flag) {	
user_info_size	u(16)
for (i=0; i < user_info_size; i++) {	
user_info_list [i]	u(v)
}	
}	
}	

<표 5-3> 제안하는 파일, 청크, 비디오 픽처 그룹별 시그널링 규격

sighted_tile_info_file {	비트수
version_info	u(8)
file_size	u(64)
for (j=0; j < file_size; j++) {	
poc_num	u(32)

info_mode	u(4)
tile_id_list_size	u(8)
for (i=0; i < list_id_list_size; i++) {	
tile_id_list [i]	u(8)
}	
if (info_mode == 4) {	
cu_id_list_size	u(16)
cu_id_list [i]	u(16)
}	
user_info_flag	u(1)
if (user_info_flag) {	
user_info_size	u(16)
for (i=0; i < user_info_size; i++) {	
user_info_list [i]	u(v)
}	
}	
}	

<표 5-4> 표 5-2와 5-3의 구문에 대한 의미론 (Semantics)

구문	의미론
version_info	시그널링 규약의 버전 정보, 부호 없는 8비트의 정보로 표현된다.
file_size	파일 사이즈, 부호 없는 64 비트의 정보로 표현된다.
poc_num	HEVC와 같은 비디오 표준에서의 POC (Picture Order Count) 정보를 의미함, 기존의 H.264 AVC 표준에서의 프레임 번호 (frame number)와 유사한 의미. 부호 없는 32 비트의 정보로 표현된다.
info_mode	본 표준에서 정의한 '정보 모드'로서 다음과 같으며, 부호 없는 4 비트의 정보로 표현된다. 0: 이전 시그널링 정보와 같음 1: 모든 타일 번호 리스트 전달 2: 연속된 타일의 시작과 끝 번호 전달 3: 좌상단 및 우하단 타일 번호 전달 4: 타일 번호 및 타일에 포함된 코딩 단위(Coding Unit: CU) 번호 리스트 전달
tile_id_list_size	타일 번호 리스트의 길이, 부호 없는 8 비트의 정보로 표현된다.
tile_id_list []	타일 번호 리스트, 각각의 타일 번호는 부호 없는 8 비트의 정보로

	표현된다.
cu_id_list_size	코딩 단위 리스트 길이, 부호 없는 16 비트의 정보로 표현된다.
cu_id_list []	코딩 단위 번호의 리스트, 각각의 코딩 단위 번호는 부호 없는 16 비트의 정보로 표현된다.
user_info_flag	추가 사용자 정보 모드의 플래그 (flag), 사용자가 추가로 전송하려는 타일 관련 정보가 있는지 여부가 부호 없는 1 비트의 정보로 표현된다. 0: 추가 사용자 정보가 없음. 1: 추가 사용자 정보가 있음.
user_info_size	추가 사용자 정보의 길이, 부호 없는 16 비트의 정보로 표현한다.
user_info_list []	추가 사용자 정보의 리스트, 각각의 추가 사용자 정보는 부호 없는 변화 가능한 (varies) 비트의 정보로 표현된다.

이상 정의된 구문과 의미론에 관한 정보들은 MPEG DASH와 같은 HTTP 기반의 영상 통신에서 각각 XML 형태로 표현이 될 수도 있다. 다음 표 5-5는 XML 형태로 (1) 정보 모드, (2) 타일 번호 리스트 길이, (3) 타일 번호 리스트의 정보를 표현한 한 예이다.

<표 5-5> XML 형태로 표현된 타일 정보 구문

```
<sighted_tile_info>
<info_mode = "3" tile_id_size = "6" tile_id_list="6, 7, 8, 9, 10, 11, 12">
</sighted_tile_info>
```

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, 마이크로소프트(MS)사의 Smooth 스트리밍(Streaming), 애플(Apple)사의 HLS (HTTP Live Streaming; HTTP 라이브 스트리밍)와 같이 미리 여러 비트레이트를 갖는 비트스트림으로 트랜스코딩 된 파일들을 사용할 경우 청크(Chunk) 단위로 높은 수준과 낮은 수준 사이를 스위치 할 수 있다.

부 록 1-1

해당사항 없음

지식재산권 확약서 정보

1-1.1 지식재산권 확약서

- 발명의 명칭: 관심 영역을 고려한 가상 현실 서비스 제공
- 권리자의 성명: 가천대학교산학협력단 (주 발명자: 류은석)
- 등록(출원) 번호: 출원: 10-2016-0125145
- 등록(출원) 연월일: 2016년 9월 28일
- 실시조건: 지적재산권을 합리적 조건하에 비차별적으로 실시하도록 하여
- 확약서 접수일: 2016년 9월 28일

부 록 1-2

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

시험인증 관련 사항

해당 사항 없음

부 록 1-3

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

본 표준의 연계(family) 표준

해당 사항 없음

부 록 1-4

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

참고 문헌

해당 사항 없음

부 록 1-5

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

영문표준 해설서

해당 사항 없음

부 록 1-6

(본 부록은 표준을 보충하기 위한 내용으로 표준의 일부는 아님)

표준의 이력

판수	채택일	표준번호	내용	담당 위원회
제1판	2016.xx.xx	이후 기재사항	이후 기재사항	이후 기재사항