

# Multi-Screen Service Forum Specification

MSS.S-Y18-003

제정일: 2018년 7월 23일

## 시선 중심의 가상현실 관심영역 정보 표현을 위한 메타데이터의 구성 요소 및 형식

Syntax and Semantics of Metadata for Information  
Representation of User Gaze based Region of Interest

제출일 : 2018. 7. 23.

제출기관 : 멀티스크린서비스포럼

제출인 : 가천대 류은석 교수

# 서 문

## 1 표준의 목적

이 표준의 목적은 가상현실에서 프로젝션을 고려하여 머리 장착형 기기가 바라보는 위치와 함께 실제로 사용자 시선이 향하는 관심영역에 대한 정보를 메타데이터로 표현함으로써, 보다 정확한 사용자 관심영역 정보를 전달하며 네트워크 환경에 따른 적응적 비디오 전송을 통해 대역폭에 최적화된 비디오 전송을 한다.

## 2 주요 내용 요약

이 표준은 머리 장착형 기기가 바라보는 위치와 사용자 시선이 향하는 관심영역에 대한 정보를 구 좌표로 표현한다. 또한 프로젝션을 고려하여 관심영역 정보를 2D 좌표와 최신 비디오 코딩 기술 국제 표준인 고효율 비디오 부호화 (High Efficiency Video Coding)의 병렬처리 기술 도구인 타일(Tile) 인덱스로 표현하며, 기술의 장점을 설명하기 위해 (1) 본 표준의 필요성, (2) 본 표준을 360 도 비디오 통신 서비스에 적용했을 때 기대되는 성능향상을 설명한다. 그리고, (3) 제안하는 표준 메타데이터의 구성요소와 형식을 기술한다.

## 3 인용 표준과의 비교

### 3.1 인용 표준과의 관련성

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

# Preface

## 1 Purpose

The purpose of this standard is to represent information about Region of Interest (ROI) of the user's gaze as metadata in consideration of projection. This standard transfers more precise user ROI and provides bandwidth-optimized video transmission through adaptive video streaming according to network environment.

## 2 Summary

This standard describes information about ROI of Head Mounted Display (HMD) and user's gaze as sphere coordinates. Considering the projection, ROI is represented by a 2D coordinate and a tile index which is a parallel processing technology tool of High Efficiency Video Coding. It consists of three sections: (1) the motivation of this standard, (2) Expected performance improvements when applying this standard to 360-degree video communication services and (3) the proposed standard metadata specifications including syntax and semantics.

## 3 Relationship to Reference Standards

The standard can use the referenced video coding standard specifications such as the High Efficiency Video Coding (HEVC) and 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 시선 중심의 가상현실 관심영역 정보 표현을 위한 메타데이터의 구성 요소 및 형식	2
5.1 사용자 시선과 프로젝션 별 타일 인덱스를 고려한 메타데이터 전송	2
5.2 표준 신호 체계 규격	3
부속서 A 본 표준의 필요성 및 확장성	6
부록 I-1 지식재산권 협약서 정보	8
I-2 시험인증 관련 사항	9
I-3 본 표준의 연계(family) 표준	10
I-4 참고 문헌	11
I-5 영문표준 해설서	12
I-6 표준의 이력	13

# 시선 중심의 가상현실 관심영역 정보 표현을 위한 메타데이터의 구성 요소 및 형식

## (Syntax and Semantics of Metadata for Information Representation of User Gaze based Region of Interest)

### 1 적용 범위

본 표준의 적용 범위는 비디오 통신에서 전달되는 메타데이터 정보를 정의하며, 360도 비디오 부호화 표준의 부가정보를 전달하는 Omnidirectional Media Application Format (OMAF)를 통해 전해질 수 있다. 또한, 본 표준의 메타데이터는 (1) 세션(Session) 정보를 전달하는 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(Descript)하는 별도의 파일로(예: XML, 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

HEVC High Efficiency Video Coding  
SHVC Scalable Extension of High Efficiency Video Coding  
ERP Equirectangular Projection  
CMP Cube Map Projection  
HMD Head Mounted Display  
MPEG Moving Picture Experts Group  
JCTVC The Joint Collaborative Team on Video Coding  
SEI Supplemental Enhancement Information  
OMAF Omnidirectional Media Application Format

DASH Dynamic Adaptive Streaming over HTTP

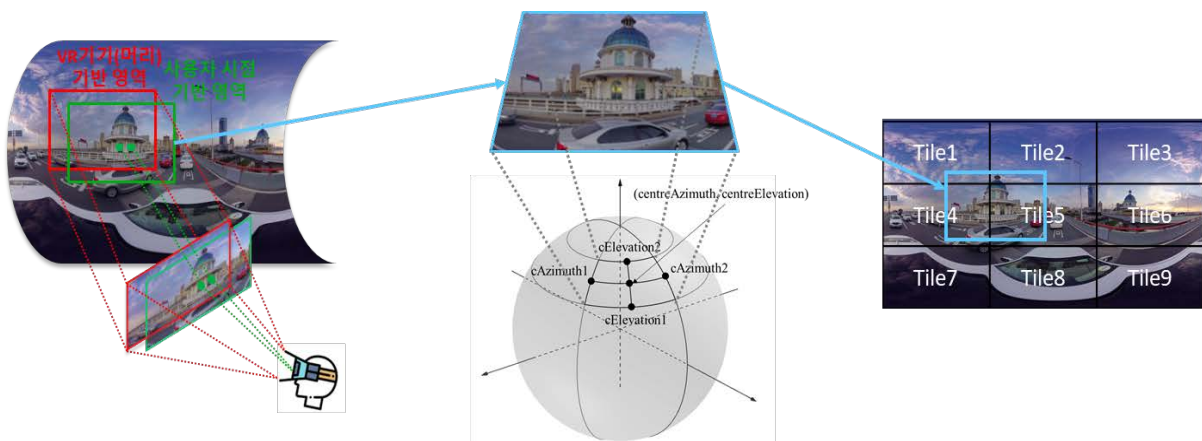
ROI Region of Interest

## 5 시선 중심의 가상현실 관심영역 정보 표현을 위한 메타데이터의 구성 요소 및 형식

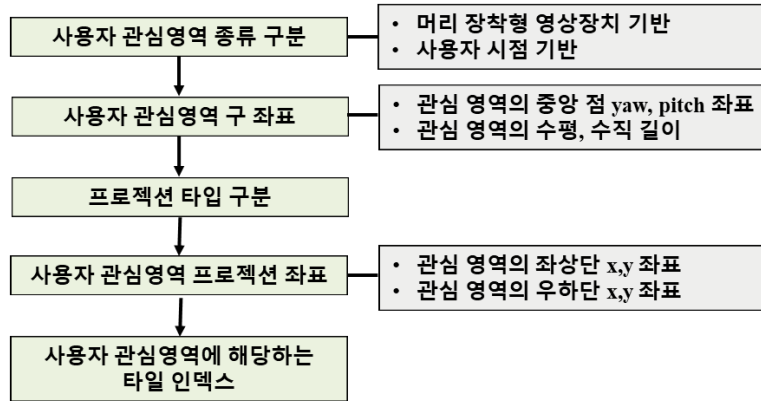
본 섹션에서는 가상현실에서 사용자 시선 중심의 관심영역 정보 표현을 위한 메타데이터 표준에 관하여 구체적인 메타데이터 규격을 설명한다.

### 5.1 사용자 시선과 프로젝션 별 타일 인덱스를 고려한 메타데이터 전송

직사각형의 프로젝션 영상은 머리 장착형 영상 장치에서 재생하기 위해 구 형태로 렌더링 된다. 따라서 (그림 5-1)과 같이 본 표준은 사용자 관심영역을 머리 장착형 영상장치 기반과 사용자 시선 기반으로 구분한 후 구 좌표를 먼저 구한 뒤 구 좌표를 프로젝션 좌표로 변환하여 사용자 관심영역에 해당하는 타일 인덱스를 구한다. 구체적으로 본 표준이 전달하는 시선 중심 관심영역의 메타데이터 순서도는 (그림 5-2)와 같다. 먼저 사용자 관심영역을 머리 장착형 영상장치 기반과 사용자 시선 기반으로 구분한다. 다음으로 선택된 관심영역의 구 좌표를 구한다. 구 좌표는 관심 영역 네 변의 중앙점들을 Yaw와 Pitch로 두 점씩 수평 수직으로 연결하여 교차하는 점을 중앙 점으로 선정한다. 여기서 관심영역의 구 좌표는 중앙 점의 yaw, pitch 좌표와 중앙점을 지나는 수평 (yaw), 수직 (pitch) 길이이다. 해당 구 좌표를 이용하여 프로젝션 종류 (예: equirectangular, cube map, etc..)에 따라 사용자 관심 영역을 맵핑 한다. 프로젝션은 직사각형 2D 이미지 이므로 관심영역의 좌표를 좌측 상단 x,y 좌표와 우측 하단 x,y 좌표로 나타낸다. 마지막으로 프로젝션의 관심영역에 해당하는 타일 인덱스를 구한다. 본 표준의 메타데이터는 머리 장착형 영상장치 기반의 관심영역 뿐만 아니라 사용자 시선 기반의 사용자 관심영역을 메타데이터로 고려하기 때문에 정밀한 관심영역 정보를 전달한다. 또한 관심영역의 구 좌표, 프로젝션 좌표, 타일 인덱스를 구분하여 사용자 환경에 따라 선택적으로 전송이 가능하며, 프로젝션을 구분하여 관심영역의 타일 인덱스를 구한다.



(그림 5-1) 사용자 관심영역의 구 좌표와 2D 프로젝션의 타일 인덱스 변환



(그림 5-2) 본 표준의 메타데이터 전달 방식 순서도

## 5.2 표준 신호 체계 규격

사용자는 기본적으로 사용자가 바라보고 있는 영역을 서버로 전달해야 한다. 따라서 본 표준에서는 필수로 사용자 관심영역의 구 좌표를 전달한다. 구 좌표를 통해 얻을 수 있는 프로젝션 좌표와 타일 인덱스는 서버와 사용자 간에 상황에 알맞게 양측에서 모두 연산이 가능하다. 따라서 본 표준에서는 사용자 관심영역의 프로젝션 좌표와 해당하는 타일 인덱스를 선택적으로 전송한다.

<표 5-1>은 H.264 AVC나 H.265 HEVC와 같은 국제 비디오 표준에서의 OMAF 구문 (Syntax)의 예를 보여주고 있다. <표 5-1>의 검은색 글씨는 기존의 표준 구문이며, 붉은색 글씨는 새로 추가되어야 할 내용이다. <표 5-2>는 이의 구문 설명을 한다.

<표 5-1> 제안하는 OMAF 구문의 예

구문
aligned(8) RegionOnViewportStruct(range_included_flag) {
unsigned int(32) viewport_id;
signed int(32) center_azimuth;
signed int(32) center_elevation;
if (range_included_flag) {
unsigned int(32) azimuth_range;
unsigned int(32) elevation_range;
}
if (projection_included_flag) {
ProjectionOnViewport(tile_included_flag);
}

}
aligned(8) ProjectionOnViewport(tile_included_flag) {
unsigned int(32) projection_id;
unsigned int(32) left_top_x;
unsigned int(32) left_top_y;
unsigned int(32) right_bottom_x;
unsigned int(32) right_bottom_y;
if (tile_included_flag) {
unsigned int(32) tile_id[];
}
}
aligned(8) RegionOnViewport() {
for (i = 0; i < num_regions; i++)
RegionOnViewportStruct (dynamic_range_flag)
}

전송하는 비디오 매 픽처마다 신호를 할 경우 아래의 <표 5-2>에 정의된 규격에 맞추어 고효율 비디오 부호화 타일 정보를 전달할 수 있다.

<표 5-2> 표 5-1의 구문에 대한 의미론 (Semantics)

구문	의미론
viewport_id	사용자 관심영역 종류 (예: 0 - 머리 장착형 영상장치, 1 - 사용자 시선)
center_azimuth	구 좌표 기준 사용자 관심영역의 중심점 yaw 값
center_elevation	구 좌표 기준 사용자 관심영역의 중심점 pitch 값
azimuth_range	구 좌표 기준 사용자 관심영역의 중심점을 지나는 yaw 수평 범위 값
elevation_range	구 좌표 기준 사용자 관심영역의 중심점을 지나는 pitch 수직 범위 값
projection_id	2D 프로젝션 종류 (예: 0 - ERP, 1 - CMP)
left_top_x	2D 프로젝션 기준 관심영역의 왼쪽 위 모서리 x 값
left_top_y	2D 프로젝션 기준 관심영역의 왼쪽 위 모서리 y 값
right_bottom_x	2D 프로젝션 기준 관심영역의 오른쪽 아래 모서리 x 좌표
right_bottom_y	2D 프로젝션 기준 관심영역의 오른쪽 아래 모서리 y 좌표
tile_id[]	관심영역에 해당하는 타일의 인덱스 배열



이상 정의된 구문과 의미론에 관한 정보들은 MPEG DASH와 같은 HTTP 기반의 영상 통신에서 각각 XML 형태로 표현이 될 수도 있다. 다음 <표 5-3>은 XML 형태로 사용자 관심영역 종류, 관심영역 좌표, 프로젝션 종류, 관심영역 프로젝션 좌표, 관심영역에 해당하는 타일의 인덱스 배열 정보를 표현한 한 예이다.

<표 5-3> XML 형태로 표현된 타일 정보 구문

```
< region_on_viewport_info >  
< viewport_id = "1" center_azimuth ="45" center_elevation ="127" projection_id = "0" left_top_x = "312"  
left_top_y = "523" right_bottom_x ="408" right_bottom_y ="845" tile_id = " 4 5 7 8">  
</ region_on_viewport_info >
```

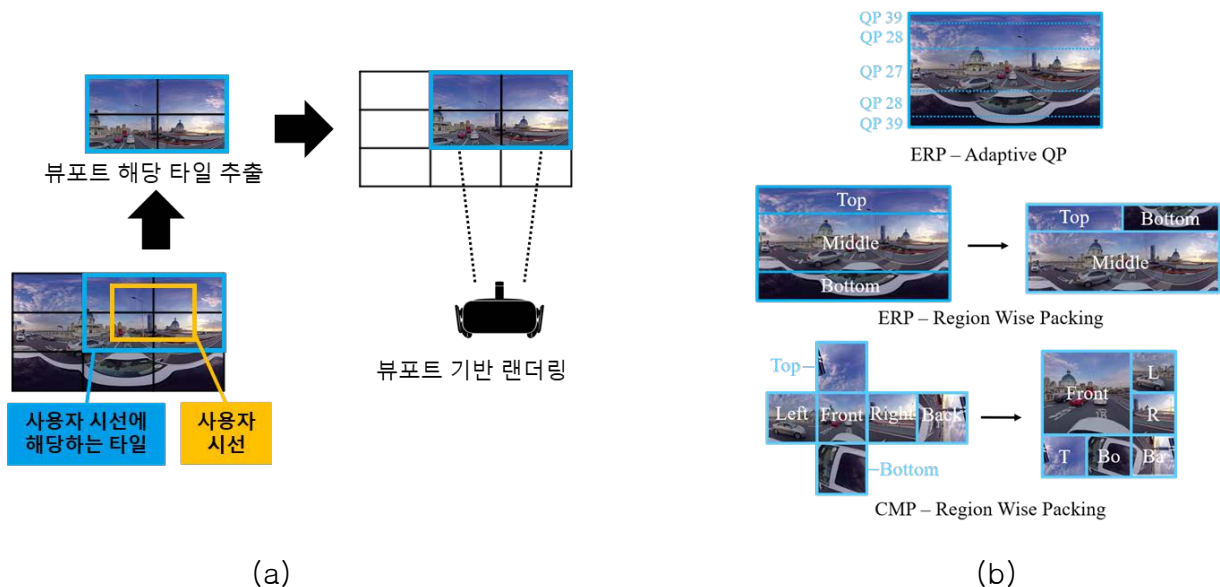
# 부속서 A

(본 부속서는 표준 내용의 일부임)

## 제목

### A.1 본 표준의 필요성

머리 장착형 영상장치 (HMD)는 사용자가 머리에 착용하여 눈앞에서 재생되고 전 방향의 360 영상을 재생해야 하기 때문에 초 고화질 영상이 요구된다. 테크니컬러 표준 문서에 따르면 사용자가 어색함 없이 가상현실을 체험하기 위해 요구되는 고 수준의 360 영상 해상도는 12K이다. 이는 Full-HD 해상도의 약 36배 크기이다. 고 수준의 360 영상은 높은 컴퓨팅 연산능력이 필요하지만 현재 출시된 머리 장착형 영상장치로 초 고화질 360 영상을 재생하기에 한계가 있다. 따라서 초 고화질 360 영상을 효율적으로 보낼 필요가 있으며, 대표적인 효율적 전송 방식으로 사용자 시선 기반 방법과 비 기반 방법이 있다. (그림 A-1)의 (b)와 같이 사용자 시선 비 기반 방식은 360 영상 전체를 보내는 대신 주요 영역 별로 QP값을 조절하거나 다운 샘플링 하여 대역폭을 낮춘다. 반면 (a)와 같이 사용자 시선 기반 방식은 360도 전체 영상 중 사용자가 바라보는 영역은 일부라는 점을 고려하여 사용자가 바라보는 영역에 해당하는 타일만 고화질로 전송한다. 사용자 시선 기반 방식은 선택적 전송으로 인해 대역폭 절감이 크며, 디코더 연산 복잡도가 낮다. 또한 사용자가 바라보지 않은 영역은 저화질로 전송하여 지연속도를 완화 시킨다. 하지만 사용자 시선을 정확히 판단하기 위해 다음의 문제점이 있다. (문제점 1) 머리 장착형 영상장치 (HMD)가 바라보는 방향은 사용자 시선이 향하는 방향과 다르다. (문제점 2) 프로젝션 종류별로 사용자 관심영역의 타일 인덱스가 다르다. 따라서, 본 표준은 사용자 시선에 대한 기준 및 사용자 시선에 해당하는 타일 인덱스를 명확히 규정하기 위한 메타데이터의 구성 요소 및 형식을 규정하고 제안한다.



(그림 A-1) (a): 사용자 시선 기반 360 비디오 전송,  
(b): 사용자 시선 비 기반 360 비디오 전송

## A.2 추가 확장적 사용

본 표준은 비트 스트림을 분할하여 전송하는 스트리밍 서비스인 MPEG DASH, 마이크로소프트(MS)사의 Smooth 스트리밍(Streaming), 애플(Apple)사의 HLS (HTTP Live Streaming; HTTP 라이브 스트리밍)에 적용 가능하다.