Multi-Screen Service Forum Specification

MSS.S-Y19-001

제정일: 2019년 8월 10일

3DoF+ 영상 처리 및 전송을 위한 메타데이터의 구성요소 및 형식

Syntax and Semantics of Metadata for 3DoF+ Video Processing and Transmission

제출일 : 2019 년 7 월 31 일

제출기관 : 멀티스크린서비스포럼

제출인 : 류은석

서 문

1 표준의 목적

이 표준의 목적은 최근에 급격히 발전하고 있는 몰입형 미디어를 위한 가상 현실(virtual reality, VR) 기술을 응용하여 여러 위치에서 촬영된 다수의 360 영상들을 전송할 때 영상 간 중복성을 제거하고 유의미한 영역 정보를 메타데이터(metadata)로 구성함으로써 다수의 360 비디오 전송 시 대역폭을 낮추고 사용자 움직임 대응을 신속하게 하여 3DoF+를 위한 저지연 전송 및 렌더링을 가능하게 하는 기술을 설명함에 있다.

2 주요 내용 요약

이 표준은 사용자 움직임 추적이 가능한 머리장착형 영상장치(head-mounted display, HMD)를 통한 다수의 360도 영상들을 전송할 때, 영상 간 중복성을 제거하고 해당 영상들에서의 유의미한 영역 정보를 추출한 뒤 기존 영상들보다 더 적은 수의 영상들로 병합하여 복원 시 필요한 정보들을 메타데이터로 기록하는 기술 및 표준 신호 체계 규격(구문과 의미론)을 기술한다.

3 인용 표준과의 비교

해당 사항 없음.

Preface

1 Purpose

The purpose of this standard is to describe technologies that enable low-latency transmission and rendering for 3DoF+ virtual reality (VR) technology to eliminate 360 videos redundancy and to organize meaningful area information into metadata, thereby reducing bandwidth and responding to user motion for multiple 360 video transmission.

2 Summary

This standard describes the technical and standard specifications (context and semantics) that extract meaningful area information from those 360 videos and record the reconstruction information as metadata in order to generate user's view screen through head-mounted devices capable of tracking user's movements.

3 Relationship to Reference Standards

목 차

3DoF+ 영상 처리 및 전송을 위한 메타데이터의 구성요소 및 형식

1 적용 범위

본 표준의 적용 범위는 비디오 통신에서의 메타데이터를 처리하는 객체를 다루며, 이는 사용자 단말, 서버, 중계 시스템 및 라우터 등을 포함한다. 또한, 본 표준의 메타데이터 신텍스(Syntax) 및 시멘틱스(Semantics)는 (1) 세션(Session) 정보를 포함하는 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: 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

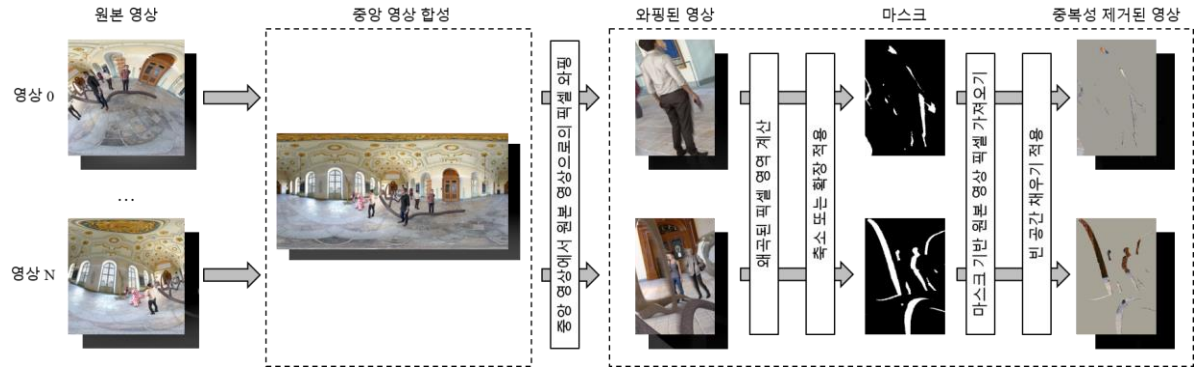
4 약어

AVC	Advanced Video Coding
DoF	Degrees of Freedom
HEVC	High Efficiency Video Coding
HTTP	Hyper Text Transfer Protocol
JCTVC	The Joint Collaborative Team on Video Coding
MPD	Media Presentation Description
MPEG	Moving Picture Experts Group
MV-HEVC	Multi-View HEVC
OMAF	Omnidirectional Media Format
VVC	Versatile Video Coding
SEI	Supplemental Enhancement Information

5 시선 기반의 360 비디오 처리를 위한 눈동자 움직임 판단 메타데이터의 구성 요소 및 형식

5.1 영상 간 압축 효율을 고려한 메타데이터 전송

3DoF+와 6DoF를 위한 시스템은 머리 장착형 영상 장치에서 사용자 시점에 대응하는 가상 영상을 생성하기 위해 다수의 360도 영상들을 전송받아 해당 시점에 렌더링한다. 본 표준은 대역폭을 절감하면서 고품질의 영상들을 전송하기 위해 (그림 5-1)과 같이 중복성 제거를 진행하는 내용을 포함한다. 먼저 기존의 360도 영상들을 사용하여 중복성을 제거할 기준이 될 영상을 생성한다. 본 표준에서는 해당 영상을 중앙 영상이라고 정의한다. 이후 중앙 영상의 픽셀들을 기존 영상들의 위치로 와핑(Warping)한다. 만약 각각의 픽셀이 중앙 영상과 기존 영상들에 의해 표현될 수 있다면 문제 없이 와핑될 것이다. 그렇지 않을 경우, (그림 5-1)의 와핑된 영상처럼 픽셀들이 왜곡되어 표현된다. 이 부분은 중앙 영상에 의하여 표현될 수 없으나 기존 영상들에 의해서만 표현될 수 있는 부분이다. 해당 영역들을 구하고 영역 축소 또는 확장을 실행하여 이진 마스크(Binary mask)의 형태로 나타낸다. 이진 마스크의 정보는 검정색 또는 하얀색으로 표기되는데, 검정색 영역은 이미 중앙 영상에 포함되어 기존 영상들에서 제외가 가능한 정보를, 하얀색 영역은 유의미한 정보를 나타낸다. 영역 확장을 실행하면 유의미한 정보의 양을 늘려 나중에 제거된 영역을 복원할 때 경계선에서의 왜곡을 줄일 수 있고, 영역 축소 실행 시 유의미한 정보의 양을 줄여 이진 마스크를 깔끔하게 정리함과 동시에 대역폭을 절감할 수 있다. 다음으로 이진 마스크에서 유의미하다고 판단된 영역의 픽셀들을 기존의 360도 영상들에서 가져온다. 마지막으로 유의미하지 않은 영역들을 정해진 값으로 채우면 영상 간 중복성이 제거된 영상이 출력된다.

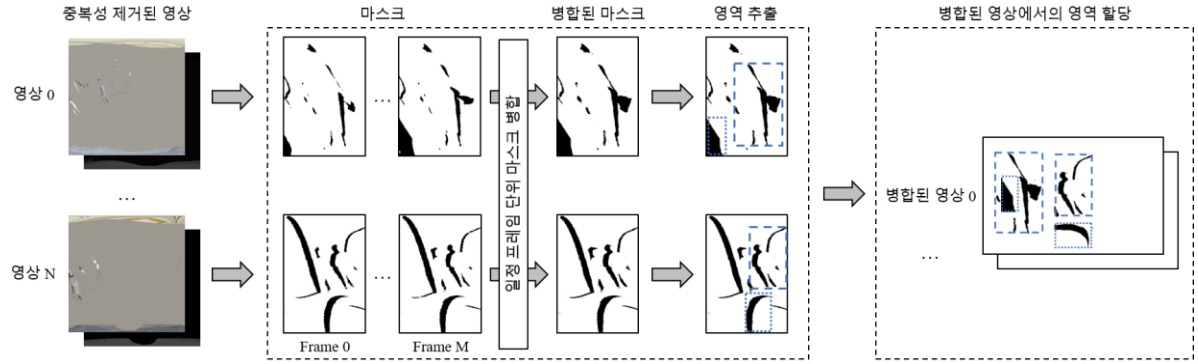


(그림 5-1) 영상 간 중복성 제거

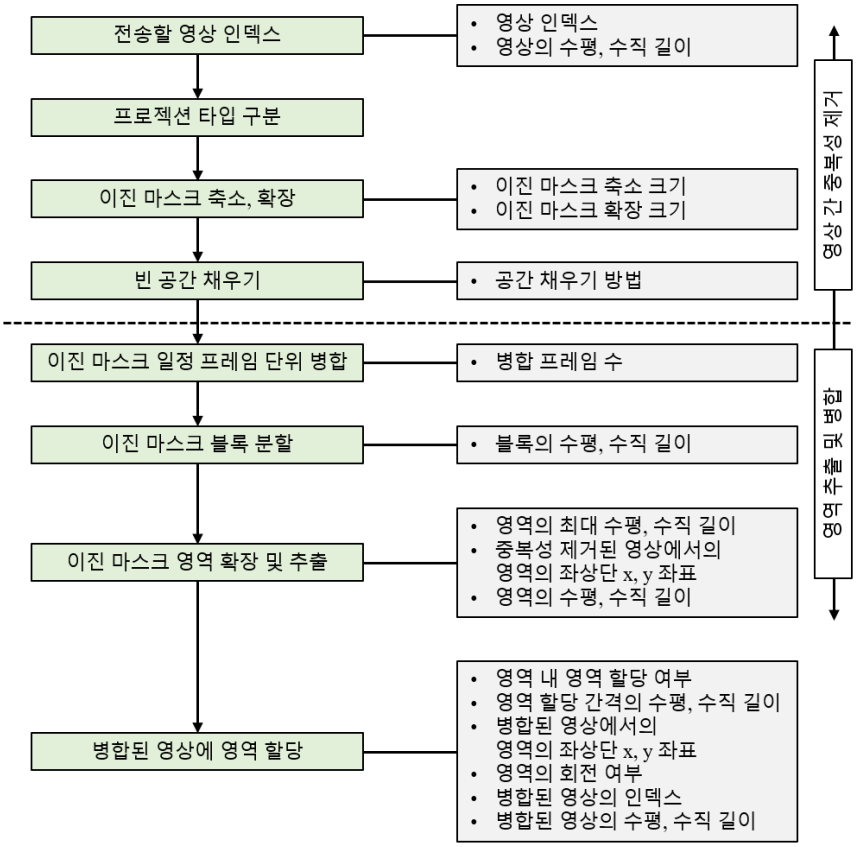
5.2 프레임 간 압축 효율과 전송할 픽셀 수를 고려한 메타데이터 전송

5.1 절에서 제안한 방법을 통해 영상 간 중복성을 제거하여 대역폭을 절감할 수 있으나, 클라이언트에 요구되는 디코더의 개수는 여전히 많고 이는 클라이언트에게 큰 부담을 야기한다. 이를 해결하기 위해 본 표준은 (그림 5-2)와 같이 중복성이 제거된 다수의 360도 영상들에서 유의미한 영역을 추출하여 기존보다 더 적은 수의 영상들로 병합하는 과정을 포함한다. 영상 간 중복성 제거 시 카메라의 이동에 따라 이진 마스크가 변하게 되는데, 유의미한 영역들을 추출할 때 한 프레임 단위로 추출한다면 매번 영역의 위치가

달라져 압축 효율이 저하될 수 있다. 이를 방지하기 위해 먼저 중복성이 제거된 360도 영상들의 유의미한 영역을 포함하는 이진 마스크를 일정 간격 단위로 병합한다. 이후 이진 마스크를 일정 크기의 블록으로 나누고 유의미한 부분을 영역 확장을 실행하여 직사각형 형태의 영역 단위로 추출하게 되는데, 해당 영역의 좌측 상단 x, y 좌표와 영역의 너비, 높이를 구한다. 마지막으로 추출된 영역들을 기존 영상들 수보다 적은 수의 영상들에 위치를 할당하여 병합한다. 본 표준은 해당 영상을 병합된 영상이라고 정의한다. 병합된 영상에서의 위치 할당 시 그림 5-2의 오른쪽 그림처럼 영역 안에 영역이 포함되게 하여 병합된 영상의 크기를 줄일 수 있다. 또한, 어떤 영역이 병합된 영상에 그대로 할당되기 어렵다고 판단될 때 영역을 회전하여 할당할 수 있다. 병합된 영상에서의 각 영역의 좌측 상단 x, y 좌표는 메타데이터로 나타내어질 수 있다. 구체적으로 본 표준이 전달하는 메타데이터 순서도는 (그림 5-3)과 같다.



(그림 5-2) 영역 추출 및 병합



(그림 5-3) 본 표준의 메타데이터 전달 방식 순서도

5.3. 표준 신호 체계 규격

서버는 중복성이 제거되어 병합된 영상들을 원래의 영상으로 복원할 수 있도록 하는 정보를 클라이언트에 전달해야 한다. 원본 영상, 병합된 영상의 개수와 인덱스, 그리고 영상의 크기에 대한 정보가 전달되어야 하며 대역폭 적응적인 다수의 360도 영상 전송을 위해 사용된 방법들에 대한 정보도 전송되어야 한다. <표 5-1>은 HEVC나 VVC와 같은 국제 비디오 표준에서의 OMAF 구문(Syntax)의 예를 보여준다.

<표 5-1> 제안하는 OMAF 구문(Syntax)의 예

구문
aligned(8) PrunedViewStruct() {
unsigned int(32) num_pruned_views;
for(int i=0; i<num_pruned_views; i++) {
unsigned int(32) pruned_view_id;
unsigned int(32) pruned_view_width;
unsigned int(32) pruned_view_height;
unsigned int(8) erosion_size;
unsigned int(8) dilation_size;
unsigned int(8) hole_filling_type;
}
}
aligned(8) PackedViewStruct() {
unsigned int(32) num_packed_views;
for(int i=0;i<num_packed_views;i++) {
unsigned int(32) packed_view_id;
unsigned int(32) packed_view_width;
unsigned int(32) packed_view_height;
unsigned int(8) intra_period;
unsigned int(32) block_width;
unsigned int(32) block_height;
unsigned int(32) max_region_width;
unsigned int(32) max_region_height;
unsigned int(32) stride_width;
unsigned int(32) stride_height;

```
        unsigned int(1) region_in_region_flag;
        RegionStruct();
    }
}
aligned(8) RegionStruct() {
    unsigned int(32) num_regions;
    for(int i=0;i<num_regions;i++) {
        unsigned int(32) region_id;
        unsigned int(32) pruned_view_id;
        unsigned int(32) region_in_pruned_view_left_top_x;
        unsigned int(32) region_in_pruned_view_left_top_y;
        unsigned int(32) region_in_packed_view_left_top_x;
        unsigned int(32) region_in_packed_view_left_top_y;
        unsigned int(32) region_width;
        unsigned int(32) region_height;
        unsigned int(1) region_rotation_flag;
    }
}
```

다음 <표 5-2>는 <표 5-1>의 구문에 대한 의미론을 설명한다.

<표 5-2> 표 5-1의 구문에 대한 의미론(Semantics)

구문	의미론
num_pruned_views	중복성이 제거된 영상들의 개수
pruned_view_id	중복성이 제거된 영상의 id
pruned_view_width	중복성이 제거된 영상의 너비
pruned_view_height	중복성이 제거된 영상의 높이
erosion_size	이진 마스크 축소 크기
dilation_size	이진 마스크 확장 크기
hole_filling_type	빈 공간 채우기 방법 (예: 0 - 중간값, 1 - 유의미한 영역들의 평균값, 2 - 보간값)
num_packed_views	병합된 영상들의 개수
packed_view_id	병합된 영상의 id
packed_view_width	병합된 영상의 너비

packed_view_height	병합된 영상의 높이
intra_period	이진 마스크 병합 프레임 수
block_width	블록의 너비
block_height	블록의 높이
max_region_width	영역의 최대 너비
max_region_height	영역의 최대 높이
stride_width	병합된 영상으로의 영역 할당 간격 너비
stride_height	병합된 영상으로의 영역 할당 간격 높이
region_in_region_flag	병합된 영상에서의 영역 내 영역 할당 여부 (예: 0 - 참, 1 - 거짓)
num_regions	영역들의 개수
region_id	영역의 id
region_in_pruned_view_left_top_x	중복성이 제거된 영상에서의 영역의 왼쪽 위 모서리 x값
region_in_pruned_view_left_top_y	중복성이 제거된 영상에서의 영역의 왼쪽 위 모서리 y값
region_in_packed_view_left_top_x	병합된 영상에서의 영역의 왼쪽 위 모서리 x값
region_in_packed_view_left_top_y	병합된 영상에서의 영역의 왼쪽 위 모서리 y값
region_width	영역의 너비
region_height	영역의 높이
region_rotation_flag	병합된 영상에서의 영역의 회전 여부 (예: 0 - 0도, 1 - 90도)

이상 정의된 구문과 의미론에 관한 정보들은 MPEG DASH와 같은 HTTP 기반의 영상 통신에서 각각 XML 형태로 표현이 될 수도 있다. 다음 표 5-3은 XML 형태로 병합된 영상 정보를 표현한 한 예이다.

<표 5-5> XML 형태로 표현된 타일 정보 구문

```
<packed_view_info>
<num_packed_views= “2”  packed_view_id= “0”  packed_view_width= “4096”
packed_view_height= “2048”  intra_period= “32”  block_width= “32”  block_height= “32”
max_region_width= “256”  max_region_height= “256”  stride_width= “32”
stride_height= “32”  region_in_region_flag= “0” >
</packed_view_info>
```

부 속 서 A

(본 부속서는 표준 내용의 일부임)

본 표준의 필요성, 배경지식, 확장적 사용

A.1 본 표준의 필요성

최근 사용자 적응적인 가상 현실 서비스에 대한 요구가 높아지면서, 사용자의 움직임에 대응하는 영상을 제공하는 use-case의 필요성이 높아지고 있다. 이를 위해 몰입형 가상 현실 서비스는 한 장면을 여러 시점에서 촬영한 영상을 전송받아 사용자 시점에 대응하는 가상 영상을 합성하여 제공하는 기술이 필요하다. 이에 MPEG-I 그룹에서는 가상 현실의 표준화 단계를 3DoF, 3DoF+, 6DoF의 3단계로 구분하고 표준화를 진행하고 있는데, 전술한 사용자 시점에 대응하는 영상을 제공하는 표준은 3DoF+, 6DoF가 해당된다. 따라서, 해당 표준은 다수의 360도 영상들을 동시에 전송할 수 있는 기술을 요구한다. 현재 영상 전송에 널리 사용되는 AVC 또는 HEVC는 개별 영상의 압축만을 전제로 하기 때문에, 동시에 여러 영상을 압축 및 전송할 때 대역폭의 한계로 인해 사용자가 만족할 만한 영상을 제공하기 어렵다는 단점이 있다. 또한 여러 비트스트림을 동시에 디코딩하여 가상 시점을 합성해야 하는 서비스 특성상 클라이언트에게도 많은 연산처리능력을 요구하는 문제가 있다. 이를 보완하기 위해 MV-HEVC 와 같은 기존 코덱에 다시점 영상을 효율적으로 압축하기 위한 기술을 추가하는 모델이 제안되었으나, 기존에 사용되던 것과 다른 디코더가 필요하고 그로 인해 하드웨어 가속 지원의 어려움과 플랫폼 파편화 등의 문제가 생겨 현재 널리 쓰이지 못하고 있다. 때문에 다수의 고품질 360도 영상을 효율적으로 압축 및 전송할 수 있는 시스템의 필요성이 지속적으로 요구되고 있다. 따라서, 본 표준은 다수의 360도 영상 간 중복성을 제거하고 유의미한 영역을 추출하여 병합한 후 전송하기 위한 메타데이터의 구성 요소 및 형식을 규정하고 제안한다.