

Carriage of Volumetric Visual Video based Coding(V3C) 국제표준 기술 동향

□ 남귀중, 김규현 / 경희대학교

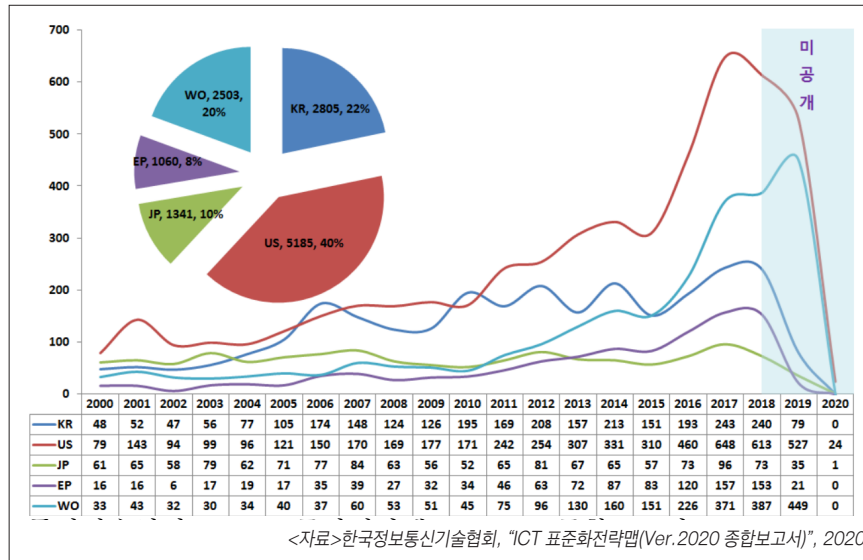
요약

최근 디바이스와 5G 통신의 비약적인 발전을 통해 가상/증강 현실 분야, 자율 주행 등 3차원 그래픽스 기술에 대한 연구가 활발하게 진행되고 있으며, 3차원 정보를 면밀하게 표현할 수 있는 포인트 클라우드와 다시점 초실감 콘텐츠가 주목받고 있다. 이와 같은 콘텐츠는 전통적인 2D 비디오 대비 많은 데이터를 사용하고 있기에, 효율적 사용을 위해서는 압축이 필수적으로 요구된다. 이에 따라 국제표준화기구인 ISO/IEC 산하 Moving Picture Expert Group(MPEG)에서는 고밀도 포인트 클라우드 및 초다시점 실감형 콘텐츠에 대한 압축 방안으로 V-PCC(Video based Point Cloud Compression) 및 MIV(MPEG Immersive Video) 기술을 표준화 중에 있으며, 또한, 압축된 데이터를 효율적으로 저장, 전송하기 위한 방안으로 Carriage of Visual Volumetric Video Coding(V3C) 표준화가 진행중에 있다. 본 고에서는 MPEG에서 진행중인 V3C 표준 기술에 대하여 살펴보고자 한다.

1. 서론

최근 디바이스와 네트워크 기술의 발전은 실감형 미디어를 위한 3D 이미지의 획득과 처리 및 전송이 가능해졌으며[1], 이에 따라 <그림 1>과 같이 전 세계적으로 실감형 미디어 분야가 활발히 연구되고 있다. 대표적인 실감형 미디어로는 포인트 클라우드와 초다시점 실감형 콘텐츠를 고려할 수 있다.

포인트 클라우드는 3차원 공간상에 색상, 반사율 등을 가지는 점들의 집합으로, 여러 대의 카메라가 설치된 스튜디오나, 라이다 등을 이용하여 획득된다. 포인트 클라우드는 3차원 공간상에서 물체를 면밀하게 표현할 수 있다는 장점을 가지고 있어 이에 따라 국제표준화기구 ISO/IEC 산하 MPEG에서는 포인트 클라우드의 압축 방법으로 Video based Point Cloud Compression(V-PCC)과 Geometry based Point Cloud Compression(G-PCC)을 표준화 중에 있다.



<그림 1> 실감방송·미디어 분야 발행국별 연도별 특허출원 동향

V-PCC는 고밀도 포인트 클라우드 객체를 압축하기 위한 방법으로 압축 과정은 다음과 같다. 3차원 포인트 클라우드 데이터를 점들의 법선 벡터를 이용하여 10개의 기준 평면으로 투영하여 2차원 패치 데이터를 생성한 뒤, 패치 데이터를 2차원 영상에 효율적으로 배치하여 2차원 이미지를 생성한다. 이 때, 점의 유무 정보를 나타내는 Occupancy, 패치의 깊이 정보를 나타내는 Geometry, 패치의 색상, 반사율과 같은 정보를 나타내는 Attribute의 3종류의 2차원 이미지가 생성될 수 있으며, 각 패치의 3차원 공간상에서의 정보를 담고 있는 아틀라스 데이터 (Atlas data)와 V3C 파라미터 정보를 저장하는 V-PCC Parameter Set(VPS) 정보가 생성된다. 생성된 아틀라스 데이터는 엔트로피 인코딩을 이용하여 압축되며, Nal unit 단위로 파일에 기록된다. 아틀라스 데이터의 Nal unit은 파라미터 정보를 저장하는 non-ACL Nal unit과, 패치 정보를 저장하는 ACL Nal unit으로 나누어진다. 또한 Occupancy, Geometry, Attribute는 HEVC와 같은 기존의 비디오 코덱을 이용하여 압축된다[2].

또한, 관심이 증대되고 있는 초다시점 실감형 콘텐츠로서, 360 비디오는 다시점 카메라를 이용하여 획득된 360도 비디오로 Texture와 Depth 영상으로 이루어져 있으며, 고정된 위치에서 방향의 전환이 가능한 3 Degrees of Freedom(3DoF)와 머리의 움직임과 같은 제한된 위치 이동이 가능하며, 방향의 전환이 가능한 3DoF+, 자유로운 이동과 방향의 전환이 가능한 6DoF로 나뉘어진다. 위치의 이동이 가능한 360 비디오의 경우 위치에 따른 View마다 Texture와 Depth 영상이 존재하므로 기존의 2D 비디오에 비해 매우 큰 크기를 가져 이를 효과적으로 활용하기 위해서는 압축이 필수적으로 요구된다. 이에 따라 ISO/IEC 산하 MPEG에서는 3DoF+ 비디오의 압축 방법으로 MPEG Immersive Video(MIV)를 표준화 중에 있다. MIV의 압축 과정은 다음과 같다. 여러 개의 View를 1-2개의 Basic view와 나머지 Additional view로 분류한 뒤, Purning 과정을 통해 Additional view 영상에서 Basic view와 중복되는 부분을 제거한다. 중복된 부분이 제거된 Additional

view 영상에서 남은 부분을 패치 형태로 추출한 후, 2차원 영상에 배치하여 2차원 이미지를 생성한다. 이 때 Texture 정보를 담고 있는 Attribute, Depth 정보를 담고 있는 Geometry, 파라미터 값들과 패치 정보를 담고 있는 Atlas data가 생성된다. 생성된 Geometry와 Attribute는 기존의 비디오 코덱을 이용하여 압축된다[3].

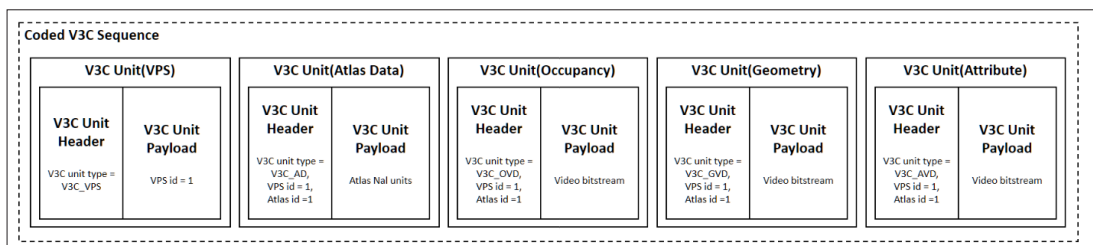
상기에서 설명한 바와 같이 V-PCC와 MIV는 압축 방식에서 매우 유사한 과정을 거치고 있기에, 해당 압축된 비트스트림을 저장하기 위한 방안으로 공통적으로 Visual Volumetric Video-based Coding(V3C) 형식으로 저장을 진행한다. V3C 비트스트림은 <그림 2>와 같이 V3C Unit header와 V3C Unit payload로 이루어진 V3C Unit 단위로 이루어져 있다. V3C Unit header는 V3C Unit type, VPS 아이디, 아틀라스 아이디 등의 정보를 포함하며, V3C Unit payload는 V3C Unit type에 따라 VPS 정보, 아틀라스 Nal unit, 인코딩된 비디오 비트스트림을 포함한다. 생성된 V3C 유닛은 멀티플렉싱되어 V3C 비트스트림을 형성하며, V3C 비트스트림은 하나 이상의 Coded V3C Sequence(CVS)로 구성된다. CVS는 VPS로 시작하며 V3C 하위 비트스트림을 운반하는 하나 이상의 V3C 장치를 포함하고 있으며, V3C 구성 요소와 관련된 각 V3C Unit을 포함한다.

상기의 V3C 비트스트림을 ISOBMFF와의 호환성을 가지며 효율적으로 저장하기 위해서 V3C 비트스트림과 재생에 필요한 메타 데이터를 패키징하는 방안이

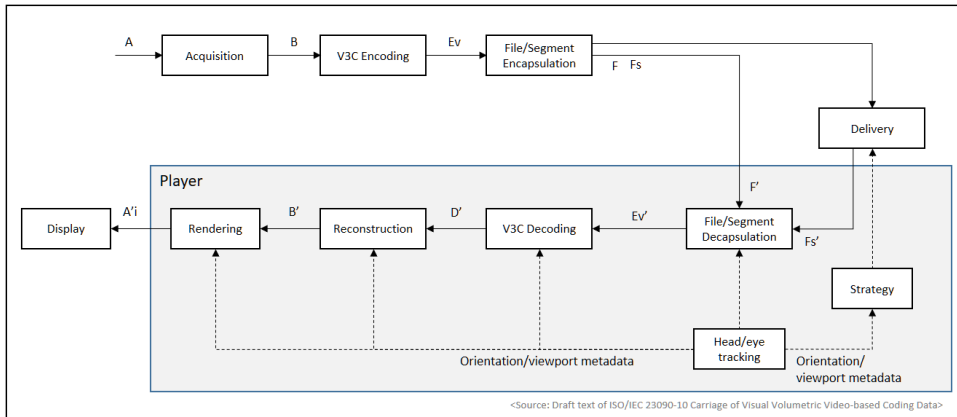
필요하며, 이를 위해 ISO/IEC 산하 MPEG WG3에서 V3C 데이터를 ISO 기반 미디어 파일 포맷(ISOBMFF)에 맞추어 패키징하는 방안인 Carriage of V3C에 대한 국제표준화가 진행 중이다. 본 고에서는 최근 FDIS 문서가 발간된 Carriage of Volumetric Visual Video-based Coding Data 표준[4]에 대해 소개하고자 한다.

II. V3C 전달 시스템

V-PCC와 MIV를 포함하는 압축 및 전달의 전체 시스템은 <그림 3>과 같은 구조를 가진다. 실제 또는 합성된 시각 장면(A)는 여러 개의 렌즈와 센서가 있는 카메라 장치에 의해 캡처(Acquisition)되어 Volumetric 데이터(포인트 클라우드 데이터 또는 초다시점 실감 콘텐츠)(B)가 생성된다. 생성된 Volumetric 데이터는 V3C 인코더(V-PCC 또는 MIV)에 의해 V3C 비트스트림(Ev)으로 인코딩(V3C Encoding) 된다. 인코딩된 V3C 비트스트림은 File/Segment Encapsulation에서 ISOBMFF를 따르는 파일로 패키징 되며, 생성된 파일은 저장이나 로컬 재생을 위한 하나의 파일(F)이나, 전송 및 스트리밍 서비스를 위한 세그먼트 파일(Fs)들로 패키징 될 수 있다. 이때 파일 또는 세그먼트 파일의 형식은 ISO/IEC 144496-12에 지정된 ISO 기반 미디어 파일 포맷(ISOBMFF)을 따른다. 이러한 과정을 통해 패키징 된 파일/세그먼트 파



<그림 2> CVS 및 V3C 유닛 구조



<그림 3> V3C 전달 시스템 구조도

일은 저장 또는 전송되어 단말에서 재생되며, 재생 과정은 다음과 같다. 저장된 파일(F') 또는 수신된 세그먼트 파일(Fs')을 파싱되어 V3C 비트스트림(Ev')으로 복원되며(File/Segment Decapsulation), V3C 디코더를 통해 디코딩된 신호(D')가 생성되고, 재구성을 통해 포인트 Volumetric 데이터(B')가 복원된다.

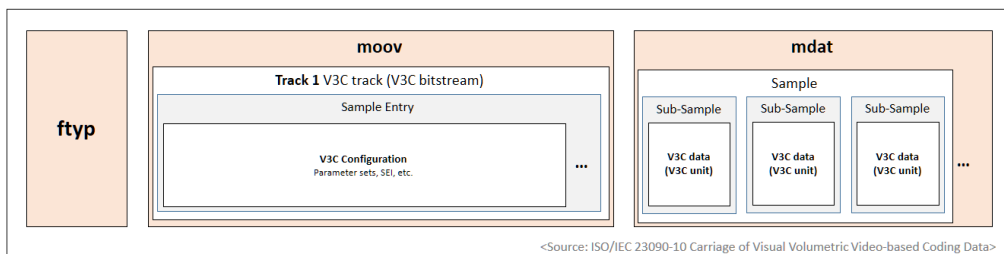
이와 같은 일련의 과정을 통해 재구성된 Volumetric 데이터는 헤드 마운트 디스플레이 등의 디스플레이 장치를 통해 화면에 표시되어 사용자에게 서비스된다. 이때 head-tracking 및 eye-tracking 기능에 의해 사용자의 시야 방향이 결정될 수 있으며, Strategy로 전달되어 시야 방향에 따라 수신될 트랙을 결정하는 뷰-포트 종속 전송이 적용될 수 있다.

III. V3C 데이터 캡슐화

V3C 데이터의 캡슐화 방안은 V3C 데이터가 하나의 트랙으로 저장되는 단일 트랙 캡슐화와 각 V3C 구성 요소가 별도의 트랙으로 저장되는 멀티 트랙 캡슐화, 비시간 V3C 데이터의 각 구성 요소가 별도의 트랙으로 저장되는 비시간 캡슐화로 나누어진다.

1. V3C 단일 트랙 캡슐화

V3C 단일 트랙 캡슐화는 V3C 데이터가 하나의 트랙으로 구성하는 캡슐화 방안으로 생성된 파일은 <그림 4>와 같은 구조를 가진다. V3C 단일 트랙 캡슐화에서 샘플



<그림 4> V3C 단일 트랙 캡슐화 파일 구조

플 엔트리에 VPS 유닛이 저장되며, 아틀라스 파라미터 정보, SEI 메시지가 저장될 수 있다. V3C 트랙의 샘플은 V3C 유닛으로 구성되며, 동일한 CVS 내의 V3C 유닛들을 서브 샘플로 포함할 수 있다. 단일 트랙 캡슐화된 V3C 데이터는 멀티 트랙 파일 생성, DASH 분할과 같은 추가 처리를 위해 사용될 수 있다.

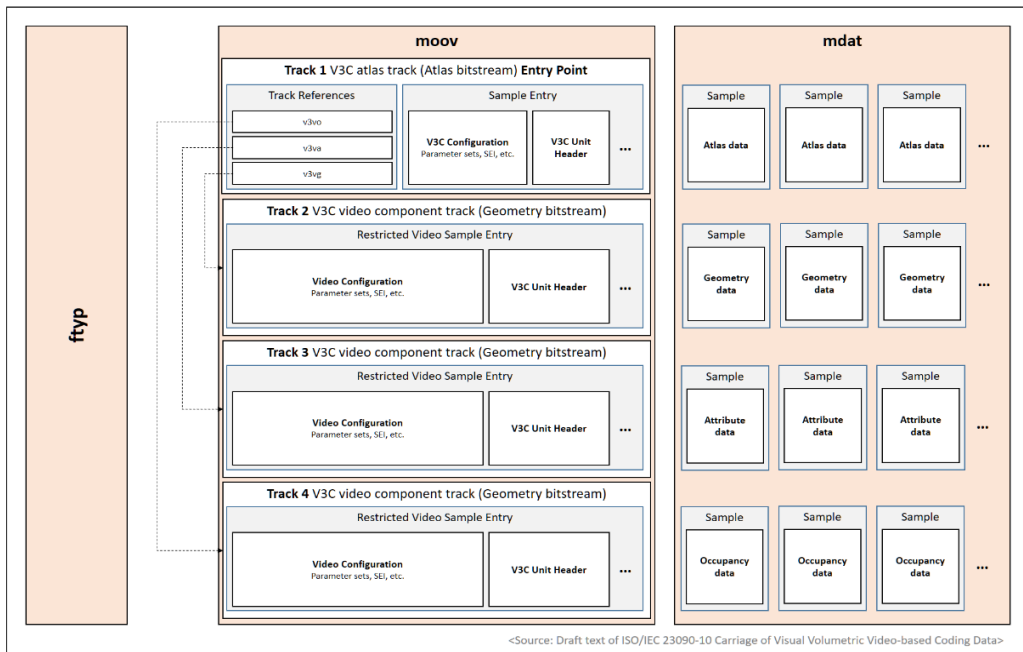
2. V3C 멀티 트랙 캡슐화

V3C 데이터의 멀티 트랙 캡슐화 방안은 V3C 비트스트림의 각 구성 요소를 각각의 트랙에 나누어 캡슐화하는 방안으로 생성된 파일은 <그림 5>와 같은 구조를 가지며 V3C 아틀라스 트랙, V3C 아틀라스 타일 트랙, V3C 비디오 구성 요소 트랙의 세 가지 유형의 트랙으로 구성되어 있다.

V3C 아틀라스 트랙에는 VPS 유닛과 아틀라스 유닛

이 저장된다. 샘플 엔트리 박스의 V3C Configuration 박스에는 VPS 유닛이 저장되며, 아틀라스 데이터 중 파라미터 정보와 SEI 메시지도 저장될 수 있다. 아틀라스 유닛의 V3C 헤더는 샘플 엔트리 내의 V3C 유닛 헤더 박스에 저장된다.

또한 V3C 비디오 구성 요소 트랙이 V3C 아틀라스 트랙 또는 V3C 아틀라스 타일 트랙과 연결됨을 나타내기 위해 ISOBMFF 트랙 참조를 활용한다. 여기서 V3C 아틀라스 트랙 또는 V3C 아틀라스 타일 트랙은 V3C 비디오 구성 요소 트랙에 대한 트랙 참조 박스를 포함한다. 아틀라스 트랙의 샘플은 아틀라스 Nal unit을 포함한다. V3C 아틀라스 타일 트랙은 포인트 클라우드 영상이 여러 개의 영역으로 나뉘어 인코딩되는 V-PCC 타일 인코딩을 사용하여 인코딩된 V3C 비트스트림을 캡슐화할 때 사용된다. V3C 아틀라스 타일 트랙이 사용될 때, V3C 아틀라스 트랙의 샘플은 디코더 초기화 정보



<그림 5> V3C 멀티 트랙 캡슐화 파일 구조

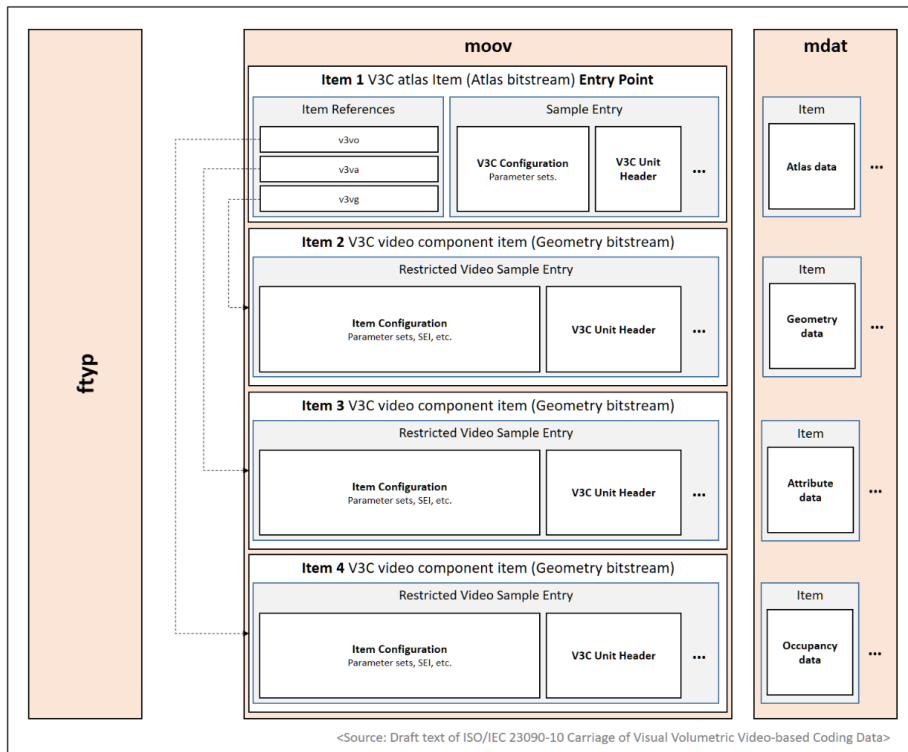
를 담고있는 VPS와 V3C 아틀라스 Non-ACL Nal unit 만을 포함하며, V3C 아틀라스 타일 트랙의 샘플은 패치 정보를 담고있는 ACL Nal unit만을 포함한다. V3C 비디오 구성 요소 트랙(V3C video component track)에는 Occupancy video unit, Geometry video unit, Attribute video unit이 각각 저장된다. V3C 비디오 트랙의 경우 샘플 엔트리 박스에 디코더 초기화 정보를 담고 있는 비디오 구성 박스를 가지고 있으며, V3C 유닛 헤더 박스에 유닛의 헤더가 저장된다. V3C 비디오 구성 요소 트랙의 샘플은 비디오 비트스트림의 액세스 유닛 단위로 존재한다. 다른 V3C 비디오 구성 요소 트랙, V3C 아틀라스 트랙 및 V3C 아틀라스 타일 트랙에서 동일한 포인트 클라우드 프레임에 기여하는 샘플은 구성 시간이 동일해야 하며, 이러한 샘플에 사용되는 아틀라

스 non-ACL 샘플(아틀라스 파라미터 정보)은 포인트 클라우드 프레임의 구성 시간 이전 또는 동일한 디코딩 시간을 가져야 한다.

3. V3C 비시간 데이터 캡슐화

V3C 비시간 데이터 캡슐화는 비시간적 포인트 클라우드 데이터에 대한 V3C 캡슐화 방안으로, 생성된 파일은 <그림 6>과 같은 구조를 가지며, V3C 아틀라스 아이템, V3C 아틀라스 타일 아이템, V3C 비디오 구성 요소 아이템의 세 가지 유형의 아이템으로 구성되어 있다.

V3C 아틀라스 아이템은 독립적으로 디코딩 가능한 아틀라스 액세스 유닛으로 아틀라스 Nal unit으로 구성되어 있다. V3C 아틀라스 타일 아이템이 사용될 때,



<그림 6> V3C 비시간적 트랙 캡슐화 파일 구조

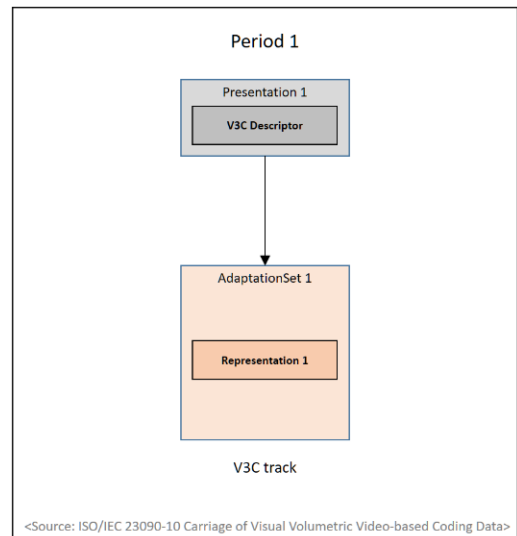
V3C 아틀라스 아이템은 non-ACL Nal unit만을 포함하며, V3C 아틀라스 타일 아이템은 ACL Nal unit만을 포함한다. 비시간 데이터 캡슐화 또한 V3C 비디오 구성 요소 아이템이 V3C 아틀라스 아이템 또는 V3C 아틀라스 타일 아이템과 연결됨을 나타내기 위해 트랙 참조를 활용하며, 여기서 V3C 아틀라스 아이템 또는 V3C 아틀라스 타일 아이템은 V3C 비디오 구성 요소 트랙에 대한 아이템 참조 박스를 포함한다.

V3C 비디오 구성 요소 아이템에는 Occupancy video unit, Geometry video unit, Attribute video unit이 각각 저장되며, 각 V3C 비디오 구성 요소 아이템은 하나의 비디오 액세스 유닛만을 포함해야 한다.

IV. V3C 기반 MPEG-DASH 서비스

MPEG-Dynamic Adaptation Streaming over Http (MPEG-DASH)[5]는 다양한 품질로 인코딩한 미디어 데이터를 서버에 저장하여, 사용자가 대역폭에 맞추어

알맞은 품질의 미디어 데이터를 요청하는 Http 기반의 표준으로 미디어 데이터의 저지연-고품질 서비스가 가능하다는 장점으로 스트리밍 분야에서 활발하게 이용되고 있다. 본 고에서 다루고 있는 Point Cloud 및 MIV 콘텐츠의 DASH 서비스를 제공하기 위해서는 V3C에서

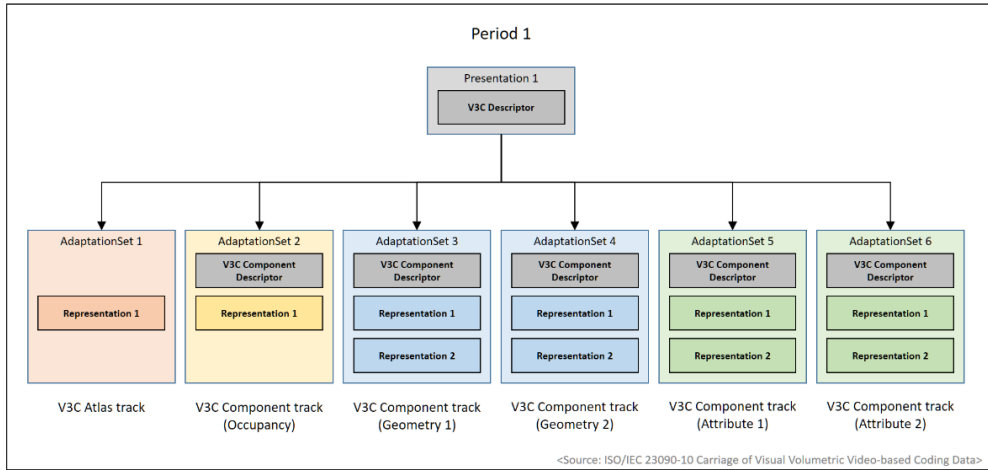


<그림 7> DASH 단일 트랙 모드 구성

```
<?xml version="1.0" encoding="UTF-8"?>
<MPD
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xmlns="urn:mpeg:dash:schema:mpd:2011"
  xsi:schemaLocation="urn:mpeg:dash:schema:mpd:2011 DASH-MPD.xsd"
  type="static"
  mediaPresentationDuration="PT3256S"
  minBufferTime="PT1.2S"
  profiles="urn:mpeg:dash:profile:isoff-on-demand:2011">
  <BaseURL>http://cdn1.example.com/</BaseURL>
  <BaseURL>http://cdn2.example.com/</BaseURL>

  <Period>
    <AdaptationSet
      mimeType="video/mp4" codecs="v3e1.L2.0.0.1, resv.vvvc.avc1.4D401E" frameRate="30">
      <SegmentList>
        <Initialization sourceURL="seg-m-init.mp4"/>
      </SegmentList>
      <Representation bandwidth="512000">
        <BaseURL>vpcc-512k.mp4</BaseURL>
      </Representation>
      <Representation bandwidth="1024000">
        <BaseURL>vpcc-1024k.mp4</BaseURL>
      </Representation>
      <Representation bandwidth="2048000">
        <BaseURL>vpcc-2048k.mp4</BaseURL>
      </Representation>
    </AdaptationSet>
  </Period>
</MPD>
```

<그림 8> DASH 단일 트랙 모드 MPD 예시



<그림 9> DASH 멀티 트랙 모드 구성

```
<?xml version="1.0" encoding="UTF-8"?>
<MPD
  xmlns="urn:mpeg:dash:schema:mpd:2011"
  xmlns:v3c="urn:mpeg:mpeg-i:v3c:2020"
  type="static"
  mediaPresentationDuration="PT10S"
  minBufferTime="PT1S"
  profiles="urn:mpeg:dash:profile:isoff-on-demand:2011">

  <Period>
    <!-- Atlas -->
    <AdaptationSet id="1" codecs="v3c1">
      <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
      <Representation>
        ...
      </Representation>
    </AdaptationSet>

    <!-- Occupancy -->
    <AdaptationSet id="2" mimeType="video/mp4" codecs="resv.vvvc.hvc1">
      <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
      <EssentialProperty schemeIdUri="urn:mpeg:mpeg-i:v3c:2020:component">
        <v3c:component component_type="occp" />
      </EssentialProperty>
      <Representation>
        ...
      </Representation>
    </AdaptationSet>

    <!-- Geometry -->
    <AdaptationSet id="3" mimeType="video/mp4" codecs="resv.vvvc.hvc1">
      <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
      <EssentialProperty schemeIdUri="urn:mpeg:mpeg-i:v3c:2020:component">
        <v3c:component component_type="geom" />
      </EssentialProperty>
      <Representation>
        ...
      </Representation>
    </AdaptationSet>

    ...

    <!-- Attribute -->
    <AdaptationSet id="5" mimeType="video/mp4" codecs="resv.vvvc.hvc1">
      <EssentialProperty schemeIdUri="urn:mpeg:dash:preselection:2016" />
      <EssentialProperty schemeIdUri="urn:mpeg:mpeg-i:v3c:2020:component">
        <v3c:component component_type="attr" />
      </EssentialProperty>
      <Representation>
        ...
      </Representation>
    </AdaptationSet>

    ...

    <!-- Preselections -->
    <Preselection id="1" tag="1" preselectionComponents="1 2 4 6" codecs="v3c1">
      <!-- V3C Descriptor -->
      <EssentialProperty schemeIdUri="urn:mpeg:mpeg-i:v3c:2020:vpc" vId="1" />
    </Preselection>

  </Period>
</MPD>
```

<Source: ISO/IEC 23090-10 Carriage of Visual Volumetric Video-based Coding Data>

<그림 10> DASH 멀티 트랙 모드 MPD 예시

정의한 Encapsulation을 기반으로 MPD에 대한 정의가 요구되며, 해당 MPD는 상기 3장에서 설명한 바와 같이, V3C ISOBMFF 파일이 단일 및 멀티 트랙 모드로 정의되어져 있기에, 이와 같은 트랙 모드에 따라 MPD 구성을 정의하여야 한다.

단일 트랙 모드의 경우는 V3C 단일 트랙 캡슐화된 파일을 이용하여 구성하며, <그림 7>과 같이 하나의 Adaptation Set으로 구성된다. 이 때, MPD는 <그림 8>과 같이 구성된다.

멀티 트랙 캡슐화의 경우 <그림 9>와 같이 각 트랙이 각각의 Adaptation Set으로 구성 가능하다. 비디오 인코딩을 하지 않는 V3C 아틀라스 트랙과 무손실 압축을 진행하는 Occupancy의 V3C 비디오 구성 요소 트랙의 경우 하나의 Representation만을 가지며, Geometry와 Attribute의 V3C 비디오 구성 요소 트랙의 경우 다양한 품질로 인코딩된 데이터가 각각의 Representation으로 구성된다. 이 때, MPD는 <그림 10>과 같이 구성된다. V3C 아틀라스 트랙의 Adaptation Set이 메인 Adaptation Set이 되며, MPD의 @Codec에 'v3c1', 'v3cg' 또는 'v3cb'를 가진다. V3C 비디오 컴포넌트 트

랙의 Adaptation Set의 경우 @Codec에는 비디오 인코딩에 사용된 코덱을 가지며, @type에는 Occupancy, Geometry, Occupancy가 각각 'occp', 'geom', 'attr'를 가진다.

V. 결론

본 고에서는 MPEG에서 표준화 중인 Carriage of Volumetric Visual Video-based Coding Data에 대해 살펴보았다. Carriage of V3C Data 표준은 V3C 데이터의 효과적인 저장을 위한 패키징 방법에 대한 표준으로, V3C 데이터의 단일 트랙 캡슐화와 멀티 트랙 캡슐화, 비시간 V3C 데이터의 캡슐화 방안에 대해 설명하고 있으며, V3C 데이터의 효과적인 전송을 위한 MPEG-DASH 구성 방안에 대해 설명하고 있다. 이에 따라 Carriage of Volumetric Visual Video based Coding은 향후 V3C 데이터를 기반으로 하는 실감형 미디어의 상용화에 크게 기여할 것으로 기대된다.

참고 문헌

- [1] F. Isgro, E. Trucco, P. Kauff and O. Schreer, "Three-dimensional image processing in the future of immersive media," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 3, pp. 288-303, March 2004
- [2] ISO/IEC FDIS 23090-5, Information technology - Video-based Point Cloud Compression
- [3] ISO/IEC DIS 23090-2, Information technology - MPEG Immersive Video
- [4] ISO/IEC FDIS 23090-10, Information technology - Carriage of Visual Volumetric Video-based Coding data
- [5] I. Sodagar, "The MPEG-DASH Standard for Multimedia Streaming Over the Internet," in IEEE MultiMedia, vol. 18, no. 4, pp. 62-67, April 2011

필자 소개



남 귀 중

- 2020년 8월 : 경희대학교 생체의공학과 공학사, 전자공학과 공학사
- 2020년 9월 ~ 현재 : 경희대학교 전자정보융합공학과 석사과정
- ORCID : <https://orcid.org/0000-0002-9502-2282>
- 주관심분야 : 영상처리, 멀티미디어 시스템



김 규 현

- 1989년 2월 : 한양대학교 전자공학과 공학사
- 1992년 9월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학석사
- 1996년 7월 : 영국 University of Newcastle upon Tyne 전기전자공학과 공학박사
- 1996년 ~ 1997년 : 영국 University of Sheffield, Research Fellow
- 1997년 ~ 2006년 : 한국전자통신연구원 대화형미디어연구팀장
- 2006년 ~ 현재 : 경희대학교 전자정보대학 교수
- ORCID : <http://orcid.org/0000-0003-1553-936X>
- 주관심분야 : 영상처리, 멀티미디어 통신, 디지털 대화형 방송