

특집논문 (Special Paper)

방송공학회논문지 제25권 제2호, 2020년 3월 (JBE Vol. 25, No. 2, March 2020)

<https://doi.org/10.5909/JBE.2020.25.2.185>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

3DoF+ 비디오의 효율적인 부호화를 위한 보호대역을 사용한 패치 패킹 기법

김 현 호^{a)}, 김 용 주^{a)}, 김 재 곤^{a)†}

A Patch Packing Method Using Guardband for Efficient 3DoF+ Video Coding

Hyun-Ho Kim^{a)}, Yong-Ju Kim^{a)}, and Jae-Gon Kim^{a)†}

요 약

MPEG-I에서는 최대 6자유도(6DoF)를 가지는 몰입형 비디오의 압축 표준화를 진행하고 있다. 360 비디오에 움직임 시차(parallax)를 추가한 것으로 정의되는 3DoF+ 비디오의 가상 공간에서, 원하는 위치와 시점의 장면을 제공하려면 3DoF+ 비디오를 구성하는 여러 뷰(view) 비디오를 사용하여 임의의 원하는 시점의 뷰를 렌더링해야 한다. MPEG-I 비주얼 그룹에서는 이러한 3DoF+ 비디오의 효율적인 부호화를 위한 표준화가 진행되고 있으며, TMIV(Test Model for Immersive Video)라는 시험모델을 개발하고 있다. 본 논문은 TMIV에서 패치(patch)를 아틀라스(atlas)에 효율적으로 패킹하여 부호화 성능을 향상시킬 수 있는 패치 패킹 방법을 제안한다. 제안 방법은 패킹되는 패치들 간에 보호대역(guardband)을 적용하여 패치 간의 거리를 뒀으로써 부호화로 인해 발생할 수 있는 아티팩트(artifact)를 줄여 최종 복원 뷰의 화질을 향상시킨다.

Abstract

MPEG-I is actively working on standardization on the immersive video coding which provides up to 6 degree of freedom (6DoF) in terms of viewpoint. In a virtual space of 3DoF+, which is defined as an extension of 360 with motion parallax, looking at the scene from another viewpoint (another position in space) requires rendering an additional viewpoint using multiple videos included in the 3DoF+ video. In the MPEG-I Visual workgroup, efficient coding methods for 3DoF+ video are being studied, and they released Test Model for Immersive Video (TMIV) recently. This paper presents a patch packing method which packs the patches into atlases efficiently for improving coding efficiency of 3DoF+ video in TMIV. The proposed method improves the reconstructed view quality with reduced coding artifacts by introducing guardbands between patches in the atlas.

Keyword : MPEG-I, 3DoF+, Immersive video, TMIV, 360 Video, VR

a) 한국항공대학교 항공전자정보공학부(School of Electronics and Information Engineering, Korea Aerospace University)

† Corresponding Author : 김재곤(Jae-Gon Kim)

E-mail: jgkim@kau.ac.kr

Tel: +82-2-300-0414

ORCID: <https://orcid.org/0000-0003-3686-4786>

※ 이 논문의 연구결과 중 일부는 한국방송·미디어공학회 “2019년 추계학술대회”에서 발표한 바 있음.

※ This work was supported by IITP grant funded by Korea Government (MSIT) (No. 2017-0-00486).

· Manuscript received January, 13, 2020; Revised February 24, 2020; Accepted February 24, 2020.

Copyright © 2020 Korean Institute of Broadcast and Media Engineers. All rights reserved.

“This is an Open-Access article distributed under the terms of the Creative Commons BY-NC-ND (<http://creativecommons.org/licenses/by-nc-nd/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited and not altered.”

I. 서론

최근 360도 비디오 및 VR(Virtual Reality) 콘텐츠가 확산되고 있으며 360도의 전방위 비디오를 넘어 사용자에게 6DoF(Degree of Freedom)까지 시점의 자유도를 제공함으로써 높은 몰입도를 제공하는 몰입형(immersive) 비디오가 주목받고 있다. 이에 발맞춰, MPEG에서는 이러한 몰입형 미디어의 압축 및 전송에 대한 표준 패키지로 MPEG-I(MPEG-Immersive) 표준화를 진행하고 있다. 즉, MPEG-I는 2016년 10월 중국 청두에서 열린 116차 회의에서 만들어진 프로젝트(“Coded Representation of Immersive Media,” ISO/IEC 23090)로 몰입형 미디어를 위한 아키텍처, 전송 포맷, 오디오/비디오 압축, 메타데이터 등을 포함한 표준 패키지를 구성하고 이에 대한 표준화 작업을 진행 중이다^{[1][2]}.

MPEG-I 비디오 그룹에서는 360도 비디오에 상체의 움직임 정도의 제한된 범위에서의 움직임 시차(motion parallax)를 제공하는 3DoF+ 비디오 부호화 표준화를 진행하고 있다. 현재 MPEG-I에서 고려되는 3DoF+ 비디오에서는 제한된 범위에서 획득 또는 생성된 10여개 이상의 360 비디오 또는 2D 비디오를 제공하고 이로부터 합성 및 렌더링(rendering)을 통하여 움직임 시차를 제공한다. 따라서 다수의 고화질(4K 또는 8K) 비디오로 구성되는 3DoF+ 비디오는 방대한 대용량 비디오가 되며 이를 효율적으로 전송하기 위한 보다 뛰어난 성능의 압축 기술이 요구된다. MPEG-I 비디오 그룹에서는 이러한 대용량의 3DoF+ 비디오를 효율적으로 압축하기 위한 표준화가 진행 중이며, 현재 WD(Working Draft)를 발간하고 표준 개발 과정에서 시

험모델인 TMIV(Test Model for Immersive Media)^[3]를 개발하고 있다.

본 논문에서는, TMIV에서 패치(patch)를 아틀라스(atlas)에 효율적으로 패킹(packing)하여 부호화 성능을 향상시킬 수 있는 패치 패킹 방법을 제안한다. 제안기법은 패킹되는 패치들 간에 보호대역(guardband)을 적용하여 패치 간의 거리를 뒀으므로, 부호화로 인해 발생할 수 있는 아티팩트(artifact)를 줄여 최종 복원 뷰의 화질을 향상시킨다. 본 논문의 구성은 다음과 같다. 2장에서는 TMIV의 패치 패킹 기법을 소개하고, 3장에서 제안기법을 제시한다. 4장의 실험결과에서는 제안기법의 성능검증과 분석 내용을 기술하고 5장에서 결론을 맺는다.

II. TMIV(Test Model for Immersive Media)

3DoF+ 비디오는 제한된 공간의 서로 다른 뷰 포인트에서 획득 및 생성된 뷰 비디오로 구성되며, 따라서 각 뷰 비디오 간에 높은 상관성이 존재한다. 3DoF+ 비디오 압축을 위한 MPEG-I의 주요 아이디어는 부호화 이전에 전처리를 통하여 3DoF+의 뷰 비디오 간에 존재하는 높은 중복성을 줄임으로써 화소를 압축(pixel rate reduction)을 하는 것이다. TMIV는 뷰 비디오 간의 중복성을 제거하는 화소율 압축의 전처리 과정과 HEVC(High Efficiency Video Coding)를 이용한 부/복호화 이후 중복성이 제거된 비디오로부터 임의의 가상시점을 합성하는 렌더링(rendering)의 후처리 과정을 포함한다.

그림 1은 TMIV 부호화기의 전체적인 구성도로 3DoF+

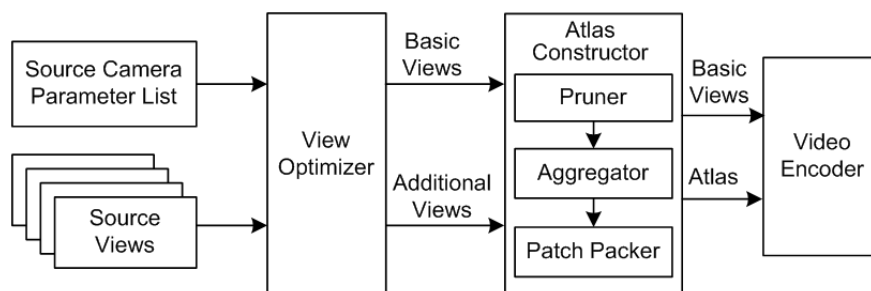


그림 1. TMIV 부호화기 구성도
Fig. 1. Architecture of TMIV encoder

비디오의 부호화 전처리 과정을 보여준다. 먼저, 제한된 공간내의 다수의 위치 위치에서 획득 또는 생성된 소스 뷰 비디오와 해당 뷰의 카메라 파라미터가 입력된다. 이들 소스 뷰들은 입력 비디오가 그대로 부호화 할 기준 뷰(Basic View)와 기준 뷰와 중복되는 영역이 제거될 부가 뷰(Additional View)로 분류된다. 다음으로, Atlas Constructor에서는 입력된 부가 뷰에서 기준 뷰와의 중복되는 영역을 제거하는 가지치기(pruning) 과정을 거친다. 가지치기는 각 뷰의 깊이 정보만을 사용하여 진행되며, 이진 마스크를 통해 각 뷰에서 중복성이 제거된 화소 위치는 0으로, 중복되지 않아 제거되지 않고 압축, 전송되어야 할 화소 위치는 1로 표현한다.

가지치기 이후, 각 뷰의 마스크에서 1 값을 가지는 화소 위치들은 클러스터링(clustering) 과정을 통해 다수의 분리

된 영역으로 구성되고, 이들 각 영역은 패치(patch)라 불리는 사각형 박스를 구성한다. 그림 2는 임의의 뷰에서 가지치기 된 결과와 패치가 생성된 예시를 보여준다. 동일한 색으로 표현된 각 영역은 하나로 클러스터링 된 영역을 나타내고 이 영역들은 각각 사각 형태의 패치를 생성하게 된다. 그림 2의 예시와 같이 사각형으로 구성된 패치는 해당 뷰에서 중복성을 제거하고 남은 영역인 유효영역 뿐만 아니라 중복으로 제거된 비유효영역까지 같이 포함하고 있음을 알 수 있다.

위 과정을 통해 각 뷰마다 생성된 패치들을 효율적으로 전송하기 위해, 생성된 모든 패치들을 아틀라스(atlas)라고 하는 하나의 프레임에 패킹한다. 그림 3은 패치가 아틀라스에 패킹 되는 과정의 예시를 보여준다. 각 패치는 아틀라스 내에서 8 픽셀씩 이동하며 최적의 패킹 위치를 찾게 되는

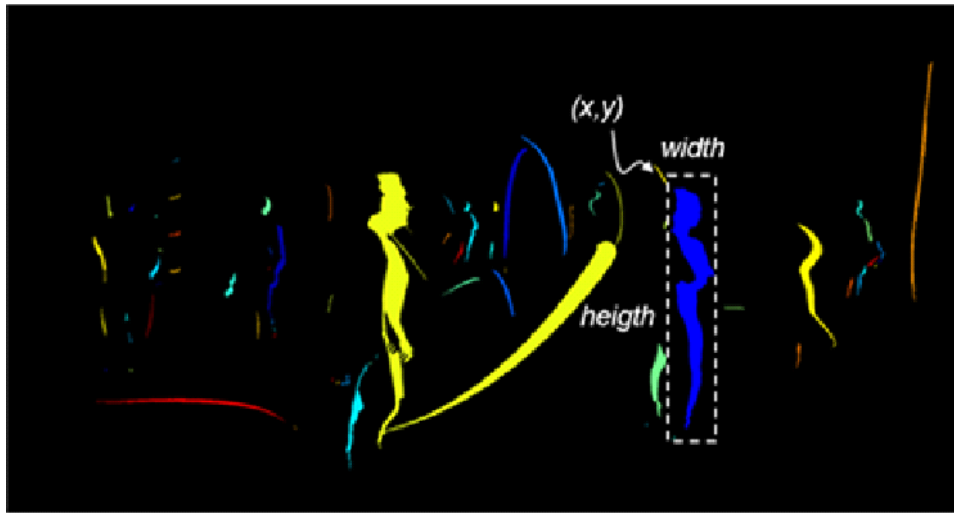


그림 2. 임의의 뷰에서의 가지치기 결과 생성된 패치의 예
Fig. 2. An example of the generated patches as the results of pruning

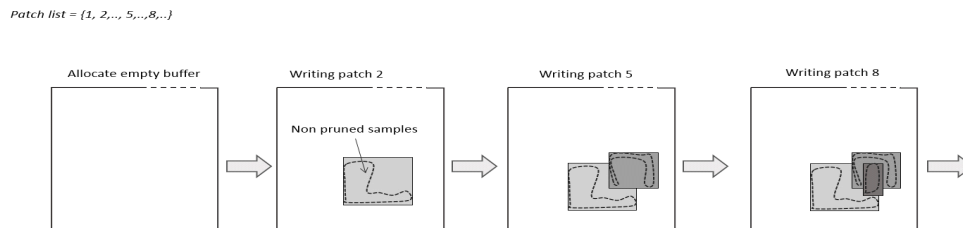


그림 3. 패치 패킹을 통한 아틀라스 생성 예
Fig. 3. An example of the generating atlas by patch packing

데, 그림 3에서와 같이 점유 공간을 최소화하기 위해 패치들 간 유효영역을 침범하지 않는 범위에서 서로 겹쳐지면서 진행된다. 또한 좀 더 유연한 패킹을 위해, 각 패치는 반시계방향으로 90도 회전하여 아틀라스에 패킹 되는 것이 가능하다. 이진 마스크를 사용하여 패킹된 패치들의 아틀라스 내 위치 및 회전 정보를 토대로, 실제 뷰에서 각 패치에 해당하는 영역의 텍스처 및 깊이 값을 아틀라스에 채워 넣음으로써 아틀라스 생성을 마무리한다.

III. 보호대역을 적용한 패치 패킹

TMIV에서 패치들이 아틀라스에 패킹되는 과정은 각 패치들이 아틀라스에서 점유하는 공간을 최소화하기 위해 패치끼리 최대한 인접하도록 배치하는 방향으로 진행된다. 이는 패치 내 비유효 영역 위에 다른 패치를 겹쳐 넣는 방식을 통해 이루어지지만, 그림 4(붉은 원)와 같이 겹쳐지는 패치가 기존 패치의 유효영역에 너무 가까이 붙어서 패킹될 경우, 불연속적 경계선이 발생함으로 인해 아틀라스의 부호화 효율 저하가 발생할 수 있다. 또한 이렇게 부호화 및 복호화된 아틀라스를 사용하여 추가 시점을 랜더링 할 경우, 불연속 경계가 나타나는 시각적 아티팩트(artifact)가 발생할 수 있다는 문제점이 있다.

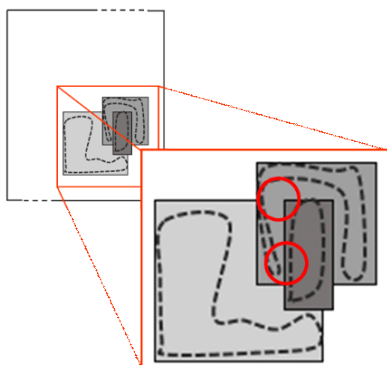


그림 4. 아틀라스 내 패킹 된 패치 예시
Fig. 4. An example of the packed patches in an atlas

이러한 문제를 해결하기 위해, 제안방법은 그림 5와 같이 패치 내 유효 영역주위에 보호대역(guardband)을 적용함으

로써 겹쳐지는 패치 간의 최소 거리를 유지하도록 한다. 각 패치는 8 픽셀 단위로 배치되어 패킹이 진행되기 때문에, 제안하는 기법에서 적용한 보호대역의 크기 또한 8 픽셀로 고정하였다. 결과적으로, 패킹되는 각 패치의 유효영역은 상하좌우로 가상의 8 픽셀이 더 존재한다는 가정 하에, 아틀라스에서 최적의 패킹 위치를 탐색하게 된다. 예를 들어, 그림 5에서 패킹하려는 패치가 그림 왼쪽의 이미 패킹되어 있는 상위 패치 내부에 패킹 가능한 영역을 탐색할 경우, 3x3 크기를 갖는 패치는 상위 패치 가까이에 붙어 패킹될 가능성이 있다. 하지만 제안방식을 적용할 경우, 패치의 크기가 5x5로 증가하게 되어 상위 패치에서 조금 더 떨어진 곳에 패킹될 수 있다. 또한 적용된 보호대역으로 인해 아틀라스 가장자리에 패치가 밀착하여 패킹되지 못한다는 단점을 보완하기 위해, 탐색하려는 위치가 아틀라스 가장자리인 경우, 위치에 따라 패치의 보호대역을 다르게 적용하였다. 예를 들어, 패치가 아틀라스의 왼쪽 가장자리를 탐색하는 중에는 보호대역을 패치의 상, 하 그리고 우측에만 적용함으로써 패치가 왼쪽에 밀착하여 패킹될 수 있도록 한다.

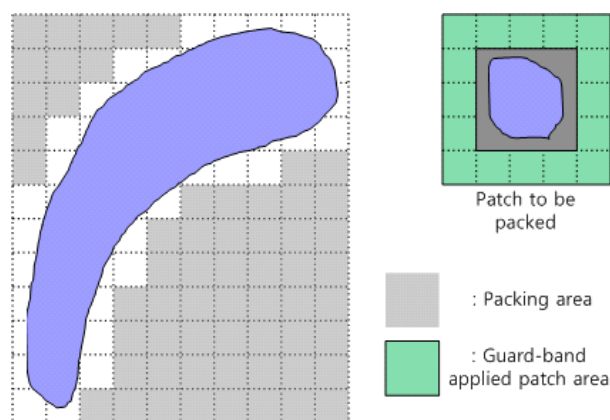


그림 5. 보호대역을 적용한 패치 패킹 예
Fig. 5. Example of patch packing with guardband

제안기법과 같이 보호대역을 적용하여 패치 패킹을 진행할 경우, 아틀라스의 점유 공간을 증가시킬 수 있지만, 패치 내 유효영역의 부호화 효율과 복원 성능을 향상시킬 수 있다는 장점이 있다. 그림 6은 TMIV를 통해 생성된 아틀라스의 예시를 보여준다. 그림 6의 위쪽은 기존 TMIV 기법으로 생성된 아틀라스이며, 아래쪽은 제안기법이 적용된 아틀라

스 생성 결과이다. 그림 6에서와 같이, 제안방식을 적용하면 패치들 간의 거리가 멀어져 아틀라스에 패킹되고, 보호대역 추가로 인한 패치의 배치도 전반적으로 변경이 되었음을 확인할 수 있다. 예를 들어, 그림의 왼쪽 상단에 있는 L 모양의 큰 패치 주변에 패킹 된 다른 패치들이 제안 방법 적용 시 해당 L 모양의 패치로부터 떨어져서 패킹 된 것을 볼 수 있다.

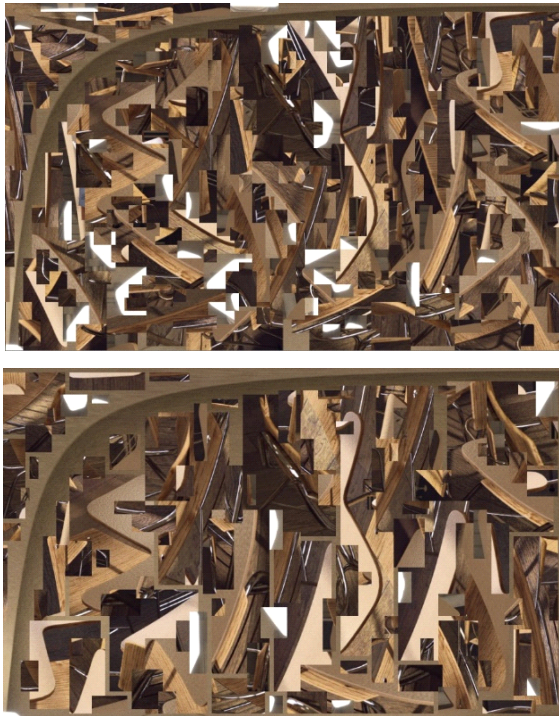


그림 6. 제안기법을 적용한 아틀라스 예(위: TMIV Anchor, 아래: 제안기법)
Fig. 6. Generated atlases with the proposed method (Above: TMIV Anchor, Below: Proposed method)

IV. 실험결과

표 1은 BD-rate 측면에서 제안기법을 적용하여 생성한 아틀라스를 포함한 3DoF+ 비디오의 단대단(end-to-end) 부호화 실험결과를 나타낸 것이다. 즉, 실험결과는 생성된 아틀라스의 부호화 성능 및 부/복호화 된 해당 아틀라스를 사용하여 입력으로 들어온 뷰의 위치를 합성해 입력 뷰와 비교한 성능이 모두 포함한 결과이다. 제안기법은 TMIV 2.0.2^[4]에 구현되었으며, MPEG-I의 몰입형 비디오 공통 테스트 조건(CTC: Common Test Condition)^[5]에 따라 실험하였다. CTC에 따라 3DoF+ 비디오의 기준 뷰와 아틀라스를 부/복호화하기 위한 핵심 코덱은 HEVC의 시험모델인 HM16.16^[6]을 사용하였다. 몰입형 비디오 공통 테스트 조건에서는 3DoF+ 비디오 특성을 고려한 화질평가를 위하여 360 비디오의 화질평가와 동일하게 WS-PSNR(Weighted Sphere PSNR)을 사용한 객관적 화질측정 방식이 사용된다. 또한, 인간의 시각 인식 특성을 고려하여 화질을 평가하는 MS-SSIM(Multi-scale Structural Similarity), VMAF (Video Multimethod Assessment Fusion) 등 다양한 척도를 사용하여 압축 성능을 측정한다.

표 1에서 녹색으로 표시된 영역은 해당 지표에서 유의미한 성능향상이 이루어졌다는 뜻이며, 반대로 붉은색으로 표시된 영역은 성능이 확연히 저하되었다는 뜻이다. 표 1의 결과로부터, 제안기법은 일부 실험영상은 VMAF와 MS-SSIM과 같은 시각 인식 특성이 고려된 화질측도에서 성능이 떨어지지만 전반적으로 높은 부호화 성능 향상을 얻을 수 있음을 확인하였다. 또한, 높은 QP 값을 사용하여 저화질로 부호화한 Low-Bitrate(Low-BR)에서의 BD-rate

표 1. 제안기법을 적용한 3DoF+ 비디오 압축 실험결과

Table 1. Experimental results on the compression of 3DoF+ videos with the proposed method

Sequence	High-BR BD rate Y-PSNR	Low-BR BD rate Y-PSNR	High-BR BD rate VMAF	Low-BR BD rate VMAF	High-BR BD rate MS-SSIM	Low-BR BD rate MS-SSIM	High-BR BD rate IV-PSNR	Low-BR BD rate IV-PSNR
ClassroomVideo	0.5%	2.9%	2.1%	3.4%	4.2%	4.6%	4.3%	3.9%
OrangeKitchen	-21.8%	-11.4%	-6.2%	0.1%	-7.0%	0.3%	-31.9%	-16.7%
TechnicolorPainter	-0.2%	1.4%	5.9%	5.0%	4.7%	4.1%	-0.2%	0.7%
IntelFrog	-6.3%	0.8%	1.7%	5.3%	0.3%	4.8%	-6.8%	0.8%
PoznanFencing	-11.1%	-6.2%	-16.1%	-10.3%	-4.7%	1.1%	-9.1%	-6.4%
Average	-7.8%	-2.5%	-2.5%	0.7%	-0.5%	2.98%	-8.7%	-3.54%

결과보다 낮은 QP 값을 사용하여 고화질로 부호화한 High-Bitrate(High-BR)에서의 BD-rate 결과가 훨씬 성능이 좋을 것을 확인할 수 있다. 이는 제안기법이 3DoF+ 비디오를 고화질로 부호화하는데 보다 더 효과적임을 확인할 수 있으며, 이러한 결과는 일반적인 미디어보다 더 고화질로 제공되어야 하는 몰입형 미디어 특성에 부합한다고 할 수 있다.

V. 결 론

본 논문은 TMIV에서 생성되는 패치에 보호대역을 추가하여 아틀라스를 생성하는 새로운 패치 패킹 기법을 제안하였다. 제안기법은 3DoF+ 비디오 부호화 성능 및 렌더링 화질 성능을 향상시킬 수 있는 기법으로 실험결과를 통해, 제안기법은 다양한 화질평가 척도에서 평균적인 BD-rate의 비트량 절감을 확인할 수 있었다. 제안한 보호대역을 적용한 패치 패킹 기법은 움직임 시차를 지원하는 전방위 비디오인 3DoF+ 비디오의 부호화 효율 향상을 위한 유용한 틀로 활용될 수 있을 것으로 기대되며, 추가 성능 개선을 위한

확장 기법의 연구를 진행할 예정이다.

참 고 문 헌 (References)

- [1] "MPEG-I Use Cases for omnidirectional 6DoF, windowed 6DoF, and 6DoF," ISO/IEC JTC1/SC29/WG11, N16768, Apr. 2017.
- [2] M. Wien, J. M. Boyce, T. Stockhammer, and W.-H. Peng, "Standardization Status of Immersive Video Coding," IEEE Jour. Emerg. Select. Topics Circuits Syst., Vol. 9, No. 1, pp. 5-17, Mar. 2019.
- [3] B. Salahieh, B. Kroon, J. Jung, M. Domański (Eds.), "Test model 2 for Immersive Video," ISO/IEC JTC1/SC29/WG11, N18577, July 2019.
- [4] J. Boyce, R. Dore, V. Vadakital, "Working Draft 2 of Immersive Video," ISO/IEC JTC1/SC29/WG11, N18576, July 2018.
- [5] J. Jung, B. Kroon, J. Boyce, "Common Test Conditions for Immersive Video," ISO/IEC JTC1/SC29/WG11, N18443, Mar. 2019.
- [6] HM reference software, [Online]. Available at: https://hevc.hhi.fraunhofer.de/svn/svn_HEVCSoftware/.
- [7] Y. Kim, H. Kim and J. Kim, "An Efficient Method of Patch Packing for 3DoF+ Video Coding," In. Proc. The Korean Institute of Broadcast and Media Engineers Fall Conference, Nov. 2019.
- [8] H. Kim, J. Kim and G. Lee, "MPEG-I 3DoF+ Video Coding Standardization," Telecommunication Technology Association, Vol 185, pp.17-25, Sept. 2019.

저 자 소 개



김 현 호

- 2018년 2월 : 한국항공대학교 항공전자정보공학과 학사
- 2020년 2월 : 한국항공대학교 항공전자정보공학과 석사
- 2020년 3월 ~ 현재 : 한국항공대학교 항공전자연구소 연구원
- ORCID : <https://orcid.org/0000-0002-8645-6142>
- 주관심분야 : 영상처리, 비디오 코덱, 360 비디오/VR



김 용 주

- 2020년 2월 : 한국항공대학교 항공전자정보공학과 학사
- ORCID : <https://orcid.org/0000-0002-8734-1515>
- 주관심분야 : 비디오 코덱, 360 비디오/VR

저 자 소 개



김 재 곤

- 1990년 2월 : 경북대학교 전자공학과 학사
- 1992년 2월 : KAIST 전기 및 전자공학과 석사
- 2005년 2월 : KAIST 전기 및 전자공학과 박사
- 1992년 3월 ~ 2007년 2월 : 한국전자통신연구원(ETRI) 선임연구원/팀장
- 2001년 9월 ~ 2002년 11월 : Columbia University 연구원
- 2007년 9월 ~ 현재 : 한국항공대학교 항공전자정보공학부 교수
- 2015년 12월 ~ 2016년 1월 : UC San Diego 방문교수
- ORCID : <https://orcid.org/0000-0003-3686-4786>
- 주관심분야 : 비디오 신호처리, 비디오 부호화 표준, Immersive Video, Deep Learning