# 딥러닝 기반의 몰입형 입체영상 압축

□ 최용호, 이진영 / 세종대학교

요 약

최근 영상처리 및 컴퓨터비전 등 많은 분야에서 딥러닝 기술이 빠르게 발전하면서 다양한 문제들을 높은 성능으로 해결하고 있다. 이에 MPEG (Moving Picture Experts Group) 표준에서도 딥러닝 기반의 미디어 기술이 활발히 제안 및 논의되고 있다. 특히, 몰입형 입체영상 압축을 위한 MPEG-I (MPEG Immersive) 표준은 메타버스 산업으로 크게 관심받고 있는 가상현실, 증강현실, 그리고 혼합현실 등에 대응하기 위해 현재 활발히 연구 중이다. 입체영상은 일반적으로 복수 시점의 컬러영상과 깊이영상으로 구성되어 있어 데이터의 양이 크기 때문에, MPEG-I 표준은 시점 간의 중복된 영역들을 제거하는 프루닝 과정을 통해 효율적인 압축을 수행한다. 하지만, 프루닝 과정에서 정반사 영역이 함께 제거되는 문제로 정확한 입체영상 복원에 한계가 있다. 본 학회지에서는 이러한 문제점을 해결하기 위하여 MPEG-I 표준에 기고된 딥러닝 기반의 정반사 영역 검출을 통한 몰입형 입체영상 압축에 대해 소개한다.

## Ⅰ. 서 론

고성능 그래픽 카드와 다양한 데이터들을 바탕으로 최근 딥러닝 기술들이 활발히 연구되고 있고, 이는 영상처리 및 컴퓨터비전 등의 분야들을 발전시키고 있다. 미디어 표준에서도 딥러닝 기술의 적용이 적극적으로 논의되고 있는데, 2020년 VVC (Versatile Video Coding) 비디오 압축 표준이 완료된 후 JVET (Joint Video Experts Team)에서는 딥러닝 기반의 비디오 압축 표준을 실험 중이다. VVC 표준은 HEVC (High Efficiency Video Coding) 표준보다 50% 이상의 압축성능 향상을 위하여 예측(Prediction), 변환(Transform), 양자화(Quantization), 그리고 필터링(Filtering) 등의 단계에서 다양한 기술들을 채택하였다[1]. 딥러닝 기반의 비디오 압축 표준에서는 각 단계마다 VVC 기술을 대체할 수 있는 딥러닝 기술이 논의되고 있다. 또 다른 표준인 VCM (Video Coding for Machines) 표준에서는 화

질 관점에서의 압축이 아닌 객체 인식 및 분할 등 비전 성능 관점에서 딥러닝 기반의 영상 정보에 대한 압축을 논의 중이다[2].

메타버스 산업으로 수요가 발생한 가상현실, 증강현실, 그리고 혼합현실 서비스에 대응하기 위하여 개발 중인 몰입형 입체영상 압축 표준 MPEG-I (MPEG Immersive)도 딥러닝 기술을 적극 고려하고 있다[3]. 기존 3D 비디오 압축 표준인 3D-AVC 및 3D-HEVC 경우에는 수평으로 배치된 카메라에서 획득한 입체영상들만을 대상으로 개발되었다[4, 5]. 따라서, 다양한 카메라 배치에서 획득한 입체영상 압축에서의 사용은 매우 비효율적이다. 입체영상은 복수 개의 카메라로부터 획득한 다양한 시점의 컬러영상과 깊이영상으로 구성되는데, 제한된 대역폭에서 많은 양의 데이터를 효율적으로 압축 및 전송하는 것이 MPEG-I의 목적이다. 이를 위하여, MPEG-I 표준은 시점 간 중복된 영역들을 제거하여 데이터의 양을 크게 줄이는 프루닝 과정을 포함하는데, 이 과정에서 정반사 영역까지 함께 제거되는 문제로 정확한 입체영상 복원에 한계가 있다. 본 학회지에서는 이러한 문제점을 해결하기 위하여 MPEG-I 표준에 기고된 딥러닝 기반의 정반사 영역 검출을 통한 몰입형 입체영상 압축에 대해 소개한다[6, 7].

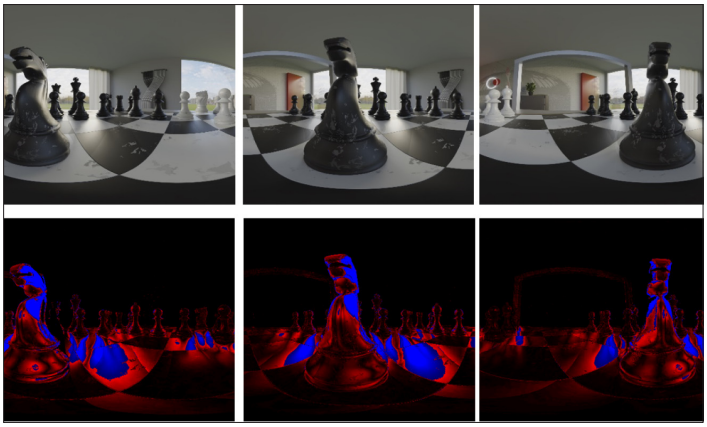
## Ⅱ. 관련 연구

MPEG-I 표준은 복수 시점의 컬러영상과 깊이영상을 효율적으로 전송하기 위하여, 시점 간 중복된 영역을 제거하는 프루닝 과정을 사용한다. 먼저, 입력 시점들을 기본 시점과 추가 시점으로 구분한다. 기본 시점에 해당하는 영상들은 별도의 과정없이 그대로 압축되지만, 추가 시점에 해당하는 영상들은 프루닝 과정을 통하여 중복된 영역들을 제거하고 나머지 영역들만 패치로 만들어 압축한다. 프루닝 과정은 시점 간 합성을 통하여 진

*<그림 1> MPEG-I 프루닝 과정에 의해 생성된 컬러영상 아틀라스*

행되고, 합성 후 두 영상 간의 차이를 계산하여 미리 정해진 임계값과 비교를 통하여 중복 영역을 결정한다. 예를 들어, 두 영상 내 픽셀값 차이가 임계값보다 작은 경우에는 중복 영역으로 결정하여 해당 영역을 제거함으로써, 압축 대상에서 제외한다. 만약 픽셀값 차이가 임계값보다 클 경우에는 해당 영역을 패치로 만든다. 마지막으로, 기본 시점의 영상들과 추가 시점에서 생성된 패치들을 아틀라스에 적절히 배치시켜, 이를 압축한다. 〈그림 1〉은 프루닝 과정 수행 후 생성된 컬러영상의 아틀라스를 보여준다.

압축된 아틀라스 정보들은 복호화 후 기본 시점의 영상들과 추가 시점의 영상들에 대한 패치들로 구분되어



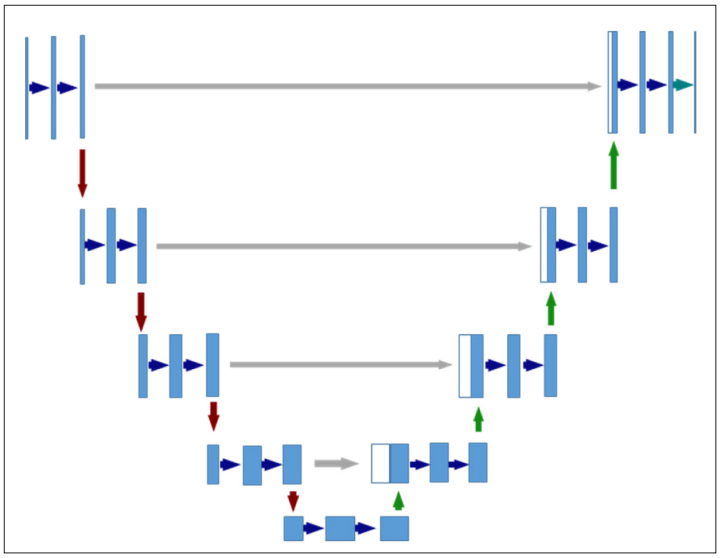
<그림 2> 정반사 영역 분석[9]

<그림 3> 정반사 프루닝 과정에 의해 생성된 컬러영상 아틀라스

입체영상 복원에 사용된다. 압축 시 프루닝 과정으로 제거된 중복 영역은 다른 시점의 영상들을 사용하여 복원된다. 입체영상 복원을 위해 임의의 시점 영상을 합성할 때는 컬러영상과 깊이영상, 그리고 카메라 파라미터 등을 활용하는 DIBR (Depth Image Based Rendering) 방식을 이용한다[8]. 하지만, 시점에 따라 다르게 보이는 정반사 영역이 프루닝 과정에 의해 제거되어 정확한 복원이 어렵고, 이는 가상현실, 증강현실, 혼합현실 등 다양한 서비스에서 몰입감을 저하시키는 원인이 되기도 한다. 이를 해결하기 위하여 〈그림 2〉의 빨간색과 파란색 영역들과 같이, 정반사 영역처럼 복원 단계에서 중요한 영향을 미치는 영역들이 프루닝 과정에서 최대한 보존될 수 있는 정반사 프루닝 과정이 MPEG-I에 기고되었다[9, 10]. 이 방법은 기존 임계값 기반의 비교 방법을 좀 더 세밀하게 수행하여, 좀 더 많은 영역을 패치로 생성시켜 아틀라스에 포함시킨다. 그 결과, 〈그림 3〉의 아틀라스 내 패치 정보가 〈그림 1〉보다 더 많아짐을 확인할 수 있다.

## Ⅲ. 딥러닝 기반의 압축 방법

정반사 프루닝 과정은 기존 MPEG-I 프루닝 과정보다 좀 더 세밀하게 임계값 비교를 수행하는 것만 추가되었기 때문에, 시점 간 픽셀값 차이가 매우 작은 경우에는 정반사 영역을 여전히 구분하지 못하는 단점이 있다[11, 12]. 이를 해결하기 위하여, 딥러닝 기술이 적용된 정반사 프루닝 과정이 MPEG-I 표준에 제안되었다[6, 7]. 제안 방법에서는 시맨틱 분할 네트워크로 유명한 UNet을 프루닝 과정에 적합하도록 최적화하여 사용하였다. 〈그림 4〉는 다운샘플링(Downsampling) 기반의 부호화 단계와 업샘플링(Upsampling) 기반의 복호화 단계로 구성된 UNet 구조를 보여준다[13]. 부호화 단계에서는 특징량(Feature)들을 점진적으로 다운샘플링시



<그림 4> 부호화 및 복호화 단계로 구성된 UNet 구조[13]

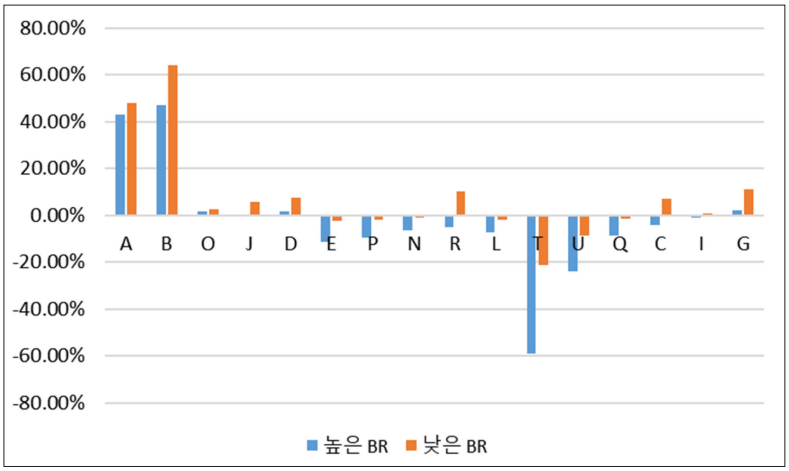
키고, 복호화 단계에서는 다시 업샘플링으로 복원한다. 그리고 부호화 단계마다 추출한 특징량들을 같은 복호화 단계로 전달하여 사용한다. 〈그림 4〉에서 색이 다른 각각의 화살표는 서로 다른 합성곱(Convolution), 풀링(Pooling) 등을 의미한다. 마지막으로, UNet 기반의 딥러닝 프루닝 과정을 통해 검출한 정반사 영역들은 기존 MPEG-I 프루닝 과정에서 생성된 패치들과 함께 아틀라스에 추가하여 압축한다.

## Ⅳ. 실험결과

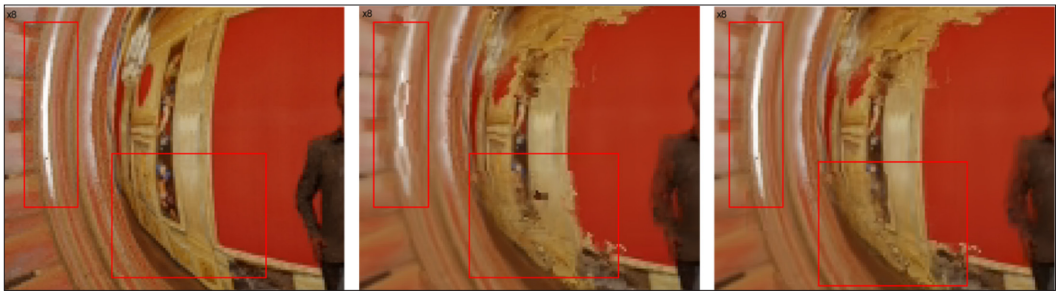
딥러닝 기반의 프루닝 과정을 포함한 몰입형 입체영상 압축 방법의 성능평가를 위하여, TMIV 11.0[14] 소프트웨어를 활용하여 CTC (Common Test Condition) [15] 기준으로 각 입체영상 시퀀스의 1장을 압축하였다. 프루닝을 위한 딥러닝 네트워크를 학습하기 위하여, 컬러영상과 정반사 영역이 표시된 마스크를 포함한 데이터를 사용하였다[16]. 이 데이터는 물체와 배경 등 정반사 영역을 다양하게 포함하고 있어 MPEG-I 입체영상 압축 실험에 적합하다. 제안 방법은 학습된 네트워크에 대해 MPEG-I 시퀀스를 입력하여 정반사 영역들을 마스크 형태로 획득한다. 그리고 해당 마스크를 기반으로 정반사 영역들을 구분하여 패치들을 생성시킨다. 마지막으로, 생성된 패치들은 아틀라스에 추가되어 압축이 수행된다. 압축성능은 동일 화질에서의 비트량(BR)으로 측정되는데, 이때 음수는 제안 방법을 통한 비트량 감소로 압축성능 향상을 의미하고, 양수는 비트량 증가로 압축성능 열화를 의미한다.

〈그림 6〉은 제안 방법의 압축성능을 나타낸다. 가로축 영문자는 MPEG-I 시퀀스를 의미하고[14], 세로축은 압축성능을 나타낸다. 동일 화질 기준으로 비트량이 감소 혹은 조금 증가를 보이지만, A와 B 시퀀스에서는 압축성능 열화가 매우 심한 것을 확인할 수 있다. 예를 들어, A와 B 시퀀스에서 압축 비트량은 40% 이상 증가한다. 이는 두 시퀀스가 대체적으로 밝기 때문에, 딥러닝 네트워크가 밝은 영역을 정반사 영역으로 잘못 검출하여, 해당 영역들이 모두 아틀라스에 포함되면서 과도한 비트량이 할당된 결과이다. 이 문제점을 해결하기 위해서는, 배경이 밝은 영상들이 학습 데이터에 많이 포함되어 있어야 한다. 〈그림 7〉은 기존 방법과 제안 방법으로 생성한 가상 시점의 영상이다. 빨간 박스와 같이, 제안 방법으로 생성한 결과영상이 원영상과 더 비슷함을 확인할 수 있다.

*<그림 5> 학습 데이터[16]*




<그림 6> 제안 방법의 압축성능



<그림 7> 원영상(좌측)과 기존(중간) 및 제안(우측) 방법으로 생성한 가상 시점의 영상

## V. 결 론

본 학회지에서 효율적인 몰입형 입체영상 압축을 위하여 딥러닝 기반의 프루닝 과정을 통해 정반사 영역을 처리하는 방법을 소개하였다. 제안 방법은 MEPG-I 입체영상 시퀀스에서 압축성능 및 주관적 화질 향상을 보여줬지만, 일부 시퀀스에서는 심각한 압축성능 열화가 확인되었다. 이는 학습 데이터와 테스트 데이터의 특징에 차이가 있어 발생한 결과로 분석되었다. 따라서, 다양한 특징을 반영한 학습 데이터가 사용된다면, 딥러닝 기술이 입체영상 압축 표준 MPEG-I에 성공적으로 적용되어 압축성능을 향상시킬 수 있을 것으로 기대된다.

## 참 고 문 헌


[1] B. Bross, Y.-K. Wang, Y. Ye, S. Liu, J. Chen, G. J. Sullivan, J.-R. Ohm, "Overview of the Versatile Video Coding (VVC) Standard and its Applications," IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 3736-3764, Oct. 2021.

[2] L. Duan, J. Liu, W. Yang, T. Huang, W. Gao, "Video Coding for Machines: A Paradigm of Collaborative Compression and Intelligent Analytics," IEEE Transactions on Image Processing, vol. 29, pp. 8680-8695, Aug. 2020.

[3] J. M. Boyce, R. Dore, A. Dziembowski, J. Fleureau, J. Jung, B. Kroon, B. Salahieh, V. K. M. Vadakital, L. Yu, "MPEG Immersive Video Coding Standard," Proceeding of the IEEE, vol. 109, no. 9, 1521-1536, Sep. 2021.

[4] J. Y. Lee, J.-L. Lin, Y.-W. Chen, Y.-L. Chang, I. Kovliga, A. Fartukov, M. Mishurovskiy, H.-C. Wey, Y.-W. Huang, S.-M. Lei, "Depth-based texture coding in AVC-compatible 3D video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 25, no. 8, pp. 1347-1361, Aug. 2015.

[5] G. Tech, Y. Chen, K. Muller, J.-R. Ohm, A. Vetro, Y.-K. Wang, "Overview of the Multiview and 3D extensions of high efficiency video coding," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 1, pp. 35-49, Jan. 2016.

[6] Y. Choi, T. V. Le, G. Bnag, J. Lee, J. Kang, J. Y. Lee, "Feature MPEG Immersive Video Coding Based on Specular Detection," ISO/IEC JTC1/SC29/WG04, m57981, Oct. 2021.

[7] Y. Choi, T. V. Le, G. Bang, J. Lee, J. Kang, J. Y. Lee, "Deep Learning Based Specular Pruning," ISO/IEC JTC1/SC29/WG04, m58997, Jan. 2022.

[8] C. Fehn, "Depth-image-based rendering (DIBR), compression, and transmission for a new approach on 3D-TV," SPIE, vol. 5291, pp. 93-104, 2004.

[9] P. R. Alface, D. Naik, V. K. M. Vadakital, J. Kerä nen, "[MPEG-I][MIV] Multiple Texture Patches per Geometry Patch," ISO/IEC JTC1/SC29/WG04, m55977, Jan. 2021.

[10] P. R. Alface, D. Naik, V. K. M. Vadakital, "MIV Exploration Experiment EE-4: multiple texture patches per geometry patch: software description and preliminary results," ISO/IEC JTC1/SC29/WG04, m55977, Apr. 2021.

[11] G. Bang, J. Lee, J. Kang, Y. Choi, J. Y. Lee, "The crosscheck report for EE4.a in Future MIV exploration experiment," ISO/IEC JTC1/SC29/WG04, m56611, Apr. 2021.

[12] G. Bang, J. Lee, J. Kang, Y. Choi, J. Y. Lee, "Results for EE3 on Future MIV," ISO/IEC JTC1/SC29/WG04, m57492, Jul. 2021.

[13] O. Ronneberger, P. Fischer, T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp 234-241, Oct. 2015.

[14] B. Salahieh, J. Jung, A. Dziembowski, "Test Model 11 for MPEG Immersive Video," ISO/IEC JTC1/SC29/WG04, N0142, Oct. 2021.

[15] J. Jung, B. Kroon, "Common Test Conditions for MPEG Immersive video," ISO/IEC JTC1/SC29/WG04, N0143, Oct. 2021.

[16] G. Fu, Q. Zhang, Q. Lin, L. Zhu, C. Xiao, "Learning to Detect Specular Highlights from Real-world Images," ACM International Conference on Multimedia pp. 1873-1881, Oct. 2020.

## 필 자 소 개

### 최 용 호

- 2021년 : 세종대학교 지능기전공학부 학사
- 2021년 ~ 현재 : 세종대학교 지능기전공학과 석사과정
- 주관심분야 : 영상처리, 비디오압축

### 이 진 영

- 2006년 : 성균관대학교 정보통신공학부 학사
- 2008년 : KAIST 전기및전자공학부 석사
- 2018년 : KAIST 전기및전자공학부 박사
- 2008년 ~ 2018년 : 삼성전자 책임연구원
- 2018년 ~ 현재 : 세종대학교 지능기전공학부 교수
- 주관심분야 : 영상처리, 비디오압축