

일반논문 (Regular Paper)

방송공학회논문지 제31권 제2호, 2026년 3월 (JBE Vol.31, No.2, March 2026)

<https://doi.org/10.5909/JBE.2026.31.2.291>

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

## 점유맵을 활용한 V-PCC 속성 영상 신경망 기반 부호화 방법

이가은<sup>a)</sup>, 이소은<sup>a)</sup>, 최해철<sup>a)†</sup>

### Occupancy Map-Guided Learned Compression for V-PCC Attribute Images

Kaeun Lee<sup>a)</sup>, Soeun Lee<sup>a)</sup>, and Haechul Choi<sup>a)†</sup>

#### 요약

본 논문에서는 기존 Video-based Point Cloud Compression (V-PCC)의 속성 영상의 부호화 효율을 개선하기 위한 신경망 기반 부호화 방법을 제안한다. V-PCC는 3차원 포인트 클라우드 데이터를 속성 영상, 기하 영상, 점유맵 등의 2차원 투영 영상으로 변환한 후, 기존의 비디오 코덱을 이용하여 이를 독립적으로 부호화한다. 속성 영상, 기하 영상, 점유맵은 높은 상관성을 갖고 있으나 V-PCC는 이 특성을 활용하고 있지 않다. 특히 점유맵은 실제 3차원 점이 존재하는 유효 영역을 표현하지만, V-PCC는 모든 화소에 대해 동등한 중요도로 부호화하기 때문에 포인트 클라우드의 부호화 측면에서 최적의 부호화 효율을 달성하기 어렵다. 이를 극복하기 위해, 본 연구에서는 Deep Contextual Video Compression을 기반으로 속성 영상을 부호화할 때 점유맵을 입력 단계와 중간 단계에서 주입하고, 유효 화소와 비유효 화소에 대해 차별적인 손실 함수로 학습시킴으로써, 비유효 화소에 대한 불필요한 비트 낭비를 줄이고 유효 화소에 대해 높은 품질로 압축하도록 유도한다.

#### Abstract

This paper proposes a neural network-based coding method to improve the coding efficiency of attribute images in conventional Video-based Point Cloud Compression (V-PCC). V-PCC converts 3D point cloud data into 2D projected representations, including attribute images, geometry images, and occupancy maps, and encodes them independently using conventional video codecs. Although attribute images, geometry images, and occupancy maps exhibit strong correlations, V-PCC does not exploit these relationships during the coding process. In particular, the occupancy map represents valid regions where actual 3D points exist; however, V-PCC assigns equal importance to all pixels during encoding, which makes it difficult to achieve optimal coding efficiency for point cloud compression. To address this limitation, this study adopts Deep Contextual Video Compression as the base codec and injects the occupancy map at both the input stage and intermediate stages of attribute image coding. Furthermore, by training the network with differentiated loss functions for valid and invalid pixels, the proposed method reduces unnecessary bit consumption for invalid regions while encouraging high-quality compression for valid regions.

Keyword : Point Cloud, Deep Learning, Video Compression

## 1. 서론

포인트 클라우드는 3차원 공간상에서 물체의 형상을 점들의 집합으로 표현한 데이터 형식으로, 각 점은 공간 좌표와 색상 정보를 포함한다. 최근 증강현실, 가상현실, 자율주행 시스템 등 3차원 응용 분야가 급격히 확대되면서 공간 정보를 표현하는 핵심 매체로서 포인트 클라우드 데이터의 중요성이 부각되고 있다. 포인트 클라우드 데이터는 수백만 개의 점으로 구성되어 높은 해상도의 공간 표현이 가능하지만, 그만큼 방대한 데이터 용량을 요구하는 한계가 존재한다. 따라서 제한된 전송 대역폭 환경에서 포인트 클라우드 데이터를 효율적으로 저장 및 스트리밍하기 위해서는 높은 압축률을 유지하면서도 우수한 복원 품질을 동시에 만족시키는 효율적인 부호화 기술의 개발이 필수적이다.

국제 표준화 기구인 ISO/IEC JTC 1/SC 29 Moving Picture Experts Group (MPEG)에서는 포인트 클라우드 데이터의 활용이 증가함에 따라 이를 효율적으로 저장 및 전송하기 위한 표준화의 필요성을 인식하여 3차원 포인트 클라우드 데이터 압축 표준인 Video-based Point Cloud Compression (V-PCC)<sup>[1]</sup>과 Geometry-based Point Cloud Compression (G-PCC)<sup>[2][3]</sup>를 제정하였다. G-PCC는 기하 정보가 먼저 부호화되고 복원된 기하를 기반으로 속성을 부호화하는 구조를 가지는 포인트 클라우드 압축 표준이다. 반면, V-PCC는 3차원 데이터를 2차원 비디오 맵으로 변환하여 압축하는 방식으로, 이 과정에서 기하 정보, 속성 정보, 점유 정보가 생성된다. 기하 정보, 속성 정보, 점유 정보는 2차원 비디오 맵 형태로 변환하여 High Efficiency Video Coding (HEVC)<sup>[4]</sup>과 같은 기존의 비디오 부호화 방식으로 압축한다. 기하 영상은 각 화소의 깊이 정보를 담고

있으며, 근층(near layer)과 원층(far layer)으로 나뉘어 생성된다. 속성 영상은 색상 정보를 2차원 평면상에 표현한 영상으로, 최종 복원된 포인트 클라우드의 시각적 품질에 직접적인 영향을 미친다. 점유맵은 이진 영상으로, 2차원 투영 평면상에서 실제 3차원 점이 존재하는 유효 영역을 나타낸다. 기하 영상, 속성 영상, 점유맵은 모두 동일한 3차원 포인트 클라우드로부터 생성되었기 때문에, 특정 화소에 대해 깊이, 색상 정보 등의 정보가 서로 연관되어 있다. 그러나 기존 V-PCC는 기하 영상, 속성 영상, 점유맵이 상관관계를 가짐에도 불구하고, 각 영상을 독립적으로 압축하여 개별적인 비트스트림을 생성한다. 이러한 방식은 속성 영상을 부호화할 때 점유맵이 제공하는 3차원 점의 실제 유효 화소에 대한 정보를 효과적으로 고려하지 못하기 때문에 최적의 부호화 성능을 달성하기 어렵다.

본 연구에서는 이러한 한계를 극복하기 위해, 점유맵을 활용한 신경망 기반의 속성 영상 부호화 방법을 제안한다. 제안 방법은 신경망 기반 비디오 코덱인 Deep Contextual Video Compression (DCVC)<sup>[5]</sup>를 기반으로 속성 영상을 부호화할 때, V-PCC에서 생성되는 점유맵을 부호화 과정에 주입함으로써 속성 영상과 점유맵 간의 상관관계를 활용한다. 구체적으로, 제안 방법은 신경망 구조에 두 가지 개선 방법을 적용한다. 첫 번째로 점유맵을 조건부 부호화 모듈의 입력 단계에 직접 결합하여 유효 화소 영역에 대한 정보를 인식할 수 있도록 한다. 두 번째로는 점유맵에서 추출된 피처(feature)를 조건부 부호화 모듈의 중간 단계에 주입하여 속성 영상의 중요한 정보를 강조한다. 또한 유효 화소와 비유효 화소에 대해 중요도를 다르게 설계한 손실 함수로 신경망을 학습시킴으로써, 비유효 화소에 대한 불필요한 비트 낭비를 줄이고 유효 화소에 대한 왜곡을 최소화하도록 유도한다.

이러한 방법을 통해 3차원 포인트 클라우드의 부호화 및 복호화 관점에서 높은 부호화 효율을 갖도록 V-PCC의 압축 성능을 개선하는 것을 목표로 한다.

## II. 관련 연구

### 1. Video-based Point Cloud Compression

국제 표준화 기구 MPEG은 포인트 클라우드 데이터의

a) 국립한밭대학교 지능미디어공학과(Hanbat University)

‡ Corresponding Author : 최해철(Haechul Choi)

E-mail: choihc@hanbat.ac.kr

Tel: +82-42-821-1149

ORCID: <https://orcid.org/0000-0002-7594-0828>

※ 이 논문은 2022년도 정부(방위사업청)의 재원으로 국방기술진흥연구소의 지원을 받아 수행된 연구임(KRIT-CT-22-047, 우주계측 지능통신망 특화연구실). 또한 2026년도 교육부 및 대전광역시 지원으로 대전 RISE센터의 지원을 받아 수행된 지역혁신중심 대학지원체계(RISE)의 결과임(2026-RISE-06-002).

· Manuscript January 28, 2026; Revised March 7, 2026; Accepted March 9, 2026.

효율적인 부호화 및 전송을 위해 2013년부터 압축 기술 논의를 시작하였다. 2014년에는 탐색 실험(Exploration Experiment)을 진행하였으며, 증강현실과 가상현실에 대한 산업계의 수요가 급증하자 2017년에 Call for Proposals를 게시하며 표준화 작업을 시작하였다. 그 결과 3차원 포인트 클라우드 데이터를 2차원 비디오 맵으로 변환하여 압축하는 V-PCC 표준이 제정되었다.

V-PCC는 3차원 포인트 클라우드 데이터를 2차원 비디오 맵으로 변환하여 압축하는 방식이다. 이 과정에서 포인트 클라우드는 기하 영상, 속성 영상, 점유맵이 생성된다. 기하 영상, 속성 영상, 점유맵은 기존의 비디오 코덱을 이용하여 압축되며, 이러한 방식은 부호화 과정에서 사용하는 코덱의 종류에 종속되지 않는 코덱 애그노스틱(agnostic) 구조를 가진다. V-PCC는 현재 HEVC와 Versatile Video Coding<sup>[6]</sup>를 사용하고 있다.

V-PCC 인코더는 크게 3차원 패치 생성, 패치 패키징, 비디오 부호화 세 단계로 구성된다. 그림 1은 V-PCC 인코더의 전체 구조를 나타낸다. 먼저 3차원 패치 생성 단계에서는 포인트 클라우드의 기하학적 특성을 기반으로 점들을 여러 개의 독립적인 패치로 분할한다. 생성된 패치는 2차원 평면으로 투영되며, 이 과정에서 기하 정보와 속성 정보가 분리되어 기록된다. 또한 3차원 점이 투영된 유효 화소 영역과

비유효 화소 영역에 대한 정보가 포함된 점유맵이 함께 생성된다. 다음으로 패치 패키징 단계에서는 생성된 패치들을 2차원 평면에 효율적으로 배치하여 비디오 압축 효율을 극대화한다. 2차원 투영 평면상에서 실제 3차원 점이 존재하는 영역은 유효 화소 영역, 그렇지 않은 영역은 비유효 화소 영역이라고 하였을 때, 각 패치는 기하 정보와 속성 정보를 포함하며 비유효 화소 영역은 배경 채우기 기법 등을 통해 보완된다. 패치 패키징 과정이 끝나면 기하 영상, 속성 영상, 점유맵이 생성되며 각 영상은 기존의 비디오 코덱을 이용하여 부호화된다<sup>[7]</sup>.

## 2. Deep Contextual Video Compression

DCVC는 수작업으로 설계된 블록 기반의 변환 및 예측 기법에 의존하는 전통적인 비디오 코덱과 달리, 신경망 기반의 구조를 채택하여 영상의 공간적 및 시간적 중복성을 End-to-End 방식으로 최적화한다. 프레임 간의 중복 제거에 한계를 가지는 현재 프레임과 이전 복원 프레임으로부터 예측된 프레임 간의 단순한 차분 연산을 사용하는 방식 대신, 조건부 부호화 방식을 활용한다. 조건부 부호화 방식은 특정 조건이나 Context 정보를 활용하여 해당 조건에 종속된 확률 분포를 기반으로 데이터를 부호화하는 방식이다.

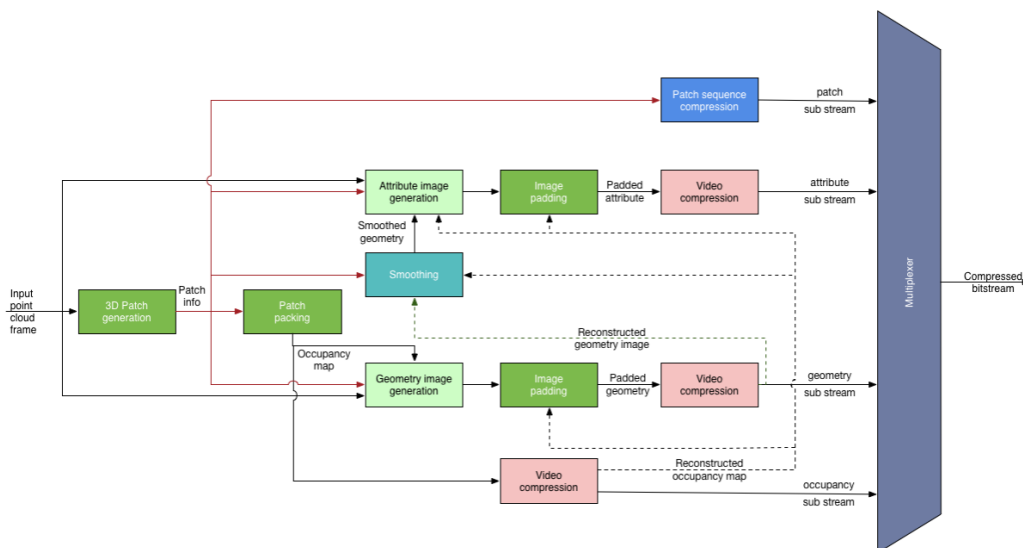


그림 1. V-PCC TMC2 참조 인코더 구조<sup>[1]</sup>  
 Fig. 1. V-PCC TMC2 reference encoder structure<sup>[1]</sup>

예측 프레임 자체를 조건으로 사용하는 경우 여전히 픽셀 도메인에 국한되어 채널 표현력이 부족하고 모델의 학습 능력이 제약되는 문제가 있다. 이를 해결하기 위해, 예측 프레임 대신 이전 복원 프레임을 피쳐 도메인으로 변환하여 움직임 추정 및 움직임 보상을 수행하여 얻은 Context를 보정하여 조건으로 활용하는 Contextual 비디오 부호화 프레임워크를 사용한다. Context는 이전 복원 프레임으로부터 추출된 고차원 피쳐 공간의 표현으로 다양한 채널을 통해 색상, 질감, 움직임 등 서로 다른 정보를 제공한다. 이러한 Context 기반의 구조는 픽셀 도메인보다 훨씬 풍부한 정보를 활용할 수 있기 때문에 고주파 영역에서 더 우수한 복원 품질을 보인다. 또한 객체의 경계나 새로운 물체가 등장하는 경우처럼 예측이 어려운 부분에서도 Context 정보를 기반으로 적응적 조건부 부호화를 수행함으로써 이를 효율적으로 압축한다. 이러한 방식을 활용하여, 현재 프레임  $x$ 는 이전 복원 프레임에서 추출한 Context  $\bar{x}$ 와 함께 Contextual 인코더를 통해 잠재 코드  $y$ 로 압축되고, 이를 양자화한  $\hat{y}$ 와 Context  $\bar{x}$ 를 이용하여 Contextual 디코더가 복원 프레임을 생성한다. 이와 같이 인코딩과 디코딩의 전체 과정은 Context  $\bar{x}$ 에 조건화되어 수행된다. Contextual 인코더는 현재 프레임  $x$ 와 Context  $\bar{x}$  간의 상관관계를 활용하여 중복 정보를 효과적으로 제거하며, Contextual 디코더 또한 동일한 Context  $\bar{x}$ 를 이용하여 양자화된 잠재 코드

로부터 복원 프레임을 생성한다. 또한 엔트로피 모델에도 Context  $\bar{x}$ 를 활용함으로써 잠재 코드가 실제로 차지하는 비트를 정확하게 예측할 수 있도록 한다.

### 3. 최신 신경망 기반 비디오 부호화

최신 신경망 기반 비디오 코덱은 피쳐 도메인에서의 시간적 Context 활용을 중심으로 발전하고 있다. 전통적인 비디오 코덱의 잔차 기반 부호화 방식은 현재 프레임과 예측 프레임 간의 단순 픽셀 차분 연산에 의존하므로 복잡한 움직임이나 고주파 성분이 포함된 영역에서 시간적 중복성을 충분히 제거하지 못하는 한계가 있다. 반면, 최신 신경망 기반 코덱은 이전 복원 프레임을 피쳐 도메인으로 변환하여 추출된 고차원 Context를 활용함으로써 부호화를 조건부 방식으로 수행하며, 이러한 방식은 채널 단위의 풍부한 피쳐 반영이 가능하다. 이러한 피쳐 기반 조건부 부호화의 대표적인 사례로는 DCVC-DC<sup>[8]</sup>, DCVC-FM<sup>[9]</sup>, DCVC-RT<sup>[10]</sup> 등이 있다. 피쳐 기반 조건부 부호화 외에도 예측의 안정성과 다양성을 확장하려는 연구들이 진행되고 있다. AlphaVC<sup>[11]</sup>는 Group of Pictures (GoP) 내의 누적 오류를 억제하기 위해 조건부 I-프레임 구조와 확률 기반 엔트로피 스킵 기법을 도입하여 성능과 계산 효율을 동시에 개선하였다. NVC-1B<sup>[12]</sup>는 대규모 파라미터를 가진 신경망 코덱

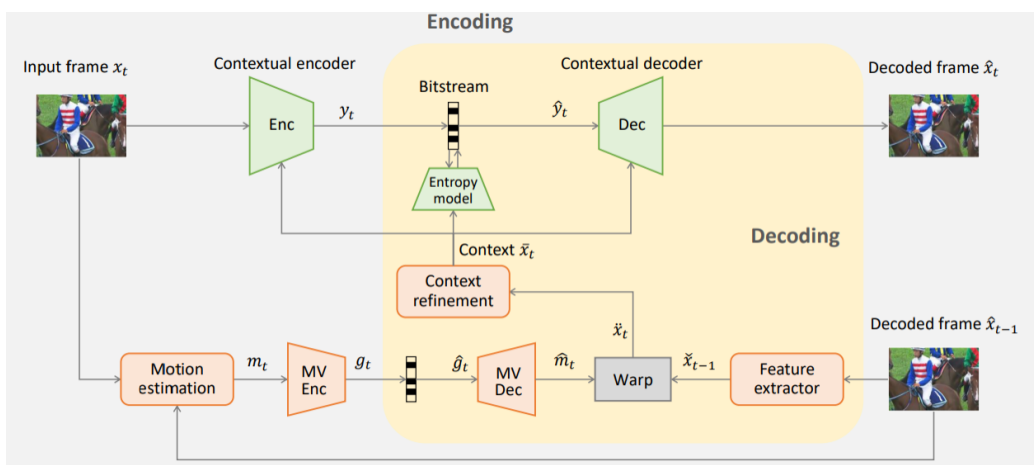


그림 2. DCVC 프레임워크<sup>[2]</sup>  
Fig. 2. The framework of DCVC<sup>[2]</sup>

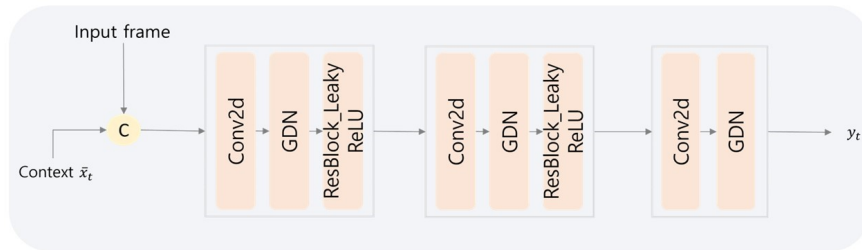


그림 3. 기존 DCVC 조건부 부호화 모듈 구조  
 Fig. 3. Architecture of the original DCVC contextual encoder

을 구축하여 CNN부터 트랜스포머에 이르기까지 다양한 아키텍처의 설계 방향을 탐구하였다. 또한 MMVC<sup>[13]</sup>는 다양한 움직임 패턴에 적응하기 위해 피쳐 도메인에서 최적의 예측 모드를 선택하는 앙상블 구조를 사용한다.

### III. 제안 방법

V-PCC 표준에서는 3차원 포인트 클라우드 데이터를 2차원 비디오 맵으로 변환한 후 이를 비디오 코덱을 통해 부호화하는 방식을 사용한다. 이 과정에서 점의 색상 정보를 가지는 속성 정보는 속성 영상으로, 3차원 공간상에서 각 화소의 존재 유무를 표시하는 점유 정보는 점유맵으로 변환된다. 그러나 속성 영상과 점유맵이 높은 상관관계가 있음에도 불구하고 V-PCC는 두 영상을 개별적인 스트림으로 분리하여 독립적으로 압축한다. 속성 영상에서 유효 화소만이 3차원 포인트 클라우드의 화질에 영향을 미침에도

불구하고, 점유맵이 속성 영상의 부호화 과정에 고려되지 않기 때문에 유효 화소 영역과 비유효 화소 영역이 동일한 중요도로 처리되어 비트가 비효율적으로 사용되는 문제가 발생한다. 속성 영상 부호화 과정에서 각 화소가 실제 3차원 공간상의 점에 대응되는지를 인지할 수 있다면, 비유효 화소 영역에 대한 중요도를 낮추고 유효 화소 영역을 중심으로 효율적인 부호화가 가능할 것이다. 이에 본 논문에서는 점유맵 정보를 속성 영상 부호화 과정에 활용함으로써 V-PCC의 속성 영상 부호화 효율을 개선하는 신경망 기반 부호화 방법을 제안한다. 제안 방법은 DCVC를 기반으로 하며 V-PCC에서 생성된 속성 영상과 점유맵이 속성 영상의 유효 화소에 대한 정보를 나타내기 때문에 현재 프레임의 확률 분포를 예측하는 부분인 조건부 부호화 모듈에 점유맵을 조건 정보로 주입하여 속성 영상을 부호화할 때 유효 화소 영역을 함께 고려하도록 하였다. 그림 3은 기존 DCVC의 조건부 부호화 모듈(Contextual Encoder) 구조이다. Conv2d는 2차원 컨볼루션 레이어를 의미하며,

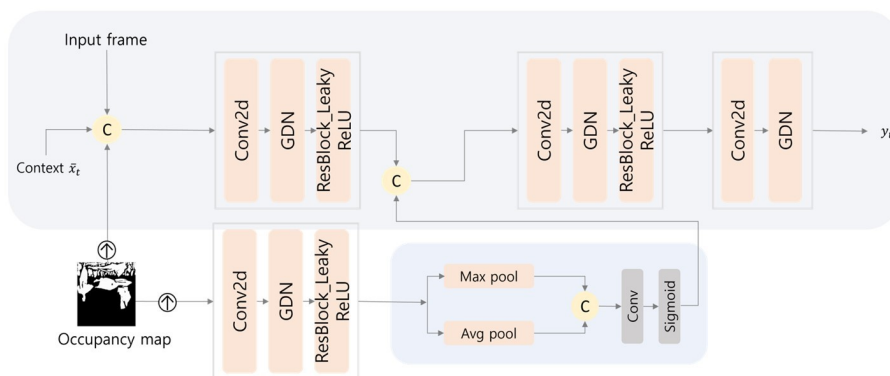


그림 4. 제안 방법을 적용한 DCVC 조건부 부호화 모듈 구조  
 Fig. 4. Architecture of the DCVC contextual encoder with the proposed method

Generalized Divisive Normalization (GDN)<sup>[14]</sup>은 피처를 정규화하는 레이어, ResBlock\_Leaky ReLU는 ResBlock 구조에 Leaky ReLU 활성화 함수를 적용한 블록을 의미한다. 기존 DCVC는 이전 복원 프레임의 Context를 조건 정보로 사용하여 현재 프레임을 부호화하는 구조를 가지지만, 본 연구에서는 점유맵을 추가적인 조건 정보로 활용하여 속성 영상의 유효 화소 영역을 함께 고려하도록 하였다. 또한 학습 과정에서는 점유맵을 활용하여 유효 화소 영역을 고려하여 왜곡을 계산하는 손실 함수를 적용함으로써 비유효 화소에 대한 불필요한 비트 낭비를 최소화하였다. 먼저 조건부 부호화 모듈의 입력 단계에서는 각 화소의 유효 화소 정보를 고려하여 특징을 추출할 수 있도록 속성 영상과 점유맵을 결합하여 입력으로 사용한다. 그러나 점유맵을 입력 단계에서 결합하는 방식만으로는 레이어의 깊이가 증가함에 따라 점유 정보가 약화될 수 있다. 따라서 점유맵 정보를 조건부 부호화 모듈의 중간 단계에서도 추가적으로 활용한다. 그림 4는 제안 방법의 조건부 부호화 모듈(Contextual Encoder) 구조이다. 조건부 부호화 모듈의 입력으로는 속성 영상, 이전 복원 프레임의 Context  $\bar{x}$ , 점유맵이 들어가고 중간 단계에서는 점유맵의 피처를 추출하여 풀링 기반의 블록을 통해 가중치 맵으로 변환한 후 속성 영상 피처와 결합한다. 구체적으로, 점유맵 피처에 대해 맥스 풀링과 평균 풀링을 각각 적용하여 서로 다른 특성의 정보를 추출한다. 맥스 풀링은 점유맵 피처 내에서 각 국소 영역에 존재하는 화소 중 가장 뚜렷하게 활성화된 값을 선택함으로써 해당 영역에서 상대적으로 중요한 정보를 강조하는 역할을 한다. 반면, 평균 풀링은 점유맵 피처의 전체적인 정보를 반영하며 맥스 풀링이 강조하지 못할 수 있는 유효 화소 영역의 정보를 보완한다. 이를 통해 조건부 부호화 모듈은 상대적으로 중요한 정보를 강조하면서 전체적인 유효 화소 영역을 고려할 수 있다.

학습을 위한 제안 방법의 손실 함수는 점유맵의 유효 화소 정보를 반영하도록 식 (1)과 식 (2)와 같이 정의된다. 왜

곡을 계산할 때, 점유맵을 활용하여 유효 화소와 비유효 화소의 상대적인 중요도를 반영한다.

여기서,  $i$ 와  $j$ 는 행과 열의 화소 좌표를 의미하고,  $d$ 는 왜곡을 나타내며,  $x$ 는 현재 프레임,  $\hat{x}$ 는 복원된 프레임,  $o$ 는 점유맵을 의미한다. 또한  $\lambda$ 는 왜곡의 가중치,  $R$ 은 인코딩 시 필요한 비트 수를 나타낸다.  $\alpha$ 는 유효 화소의 가중치로 0 이상 1 이하의 값을 가지며,  $1 - o_{i,j}$  부분이 비유효 화소 영역을 나타낸다고 가정하였을 때,  $\alpha$  값이 1에 가까울수록 유효 화소 영역에 대한 왜곡이 더 강조된다.  $\alpha$ 를 1로 설정하면 유효 화소의 왜곡만 고려하고 비유효 화소 영역의 왜곡을 무시할 수 있다. 본 실험에서는  $\alpha$  값을 1로 설정하였다. 따라서 본 IV장 실험에서의 왜곡은 유효 화소 영역에 대해서만 Mean Squared Error (MSE)를 계산하여 손실값에 영향을 주도록 하였다. 이를 통해 비유효 화소 영역의 왜곡을 의도적으로 손실 계산에서 제외하여 비트 낭비를 최소화하도록 하였다. 이는 조건부 부호화 모듈에서 점유맵을 활용하는 설계와 결합하여 속성 영상 부호화 효율을 효과적으로 개선한다.

## IV. 실험 및 분석

### 1. 실험

본 실험에서는 제안 방법의 성능을 검증하기 위해 V-PCC의 참조 소프트웨어인 TMC2 v21<sup>[15]</sup>을 활용하였으며, DCVC에 제안 방법을 적용하여 실험을 수행하였다. 제안 방법은 DCVC를 Vimeo-90k septuplet 데이터셋을 사용하여 초기 학습을 진행한 후, 포인트 클라우드 데이터에 최적화하기 위해 파인튜닝하였다. 파인튜닝을 위한 데이터셋으로는 Video-based Dynamic Mesh Coding (V-DMC)<sup>[16]</sup> Common Test Condition (CTC)<sup>[17]</sup>에 정의된 데이터셋 중 Mitch, Thomas, Football, Levi 시퀀스와 AI-based PCC (AI-PCC)

$$d(x, \hat{x}, o) = \frac{1}{N} \sum_{i,j} (\alpha \cdot o_{i,j} \cdot (x_{i,j} - \hat{x}_{i,j})^2 + (1 - \alpha)(1 - o_{i,j})(x_{i,j} - \hat{x}_{i,j})^2) \quad (1)$$

$$Loss = \lambda d(x_t, \hat{x}_t, o_t) + R \quad (2)$$

CTC<sup>[18]</sup>에 정의된 데이터세트 중 Model, Thaidancer 시퀀스, 그리고 MPEG에서 지원하는 OwlII Dynamic Human Textured Mesh Sequence Dataset의 Exercise 시퀀스를 사용하였다. 테스트 데이터세트로는 V-PCC CTC<sup>[19]</sup>에 정의된 데이터세트를 사용하였다. 모든 실험 환경은 Random Access (RA) 부호화 환경에서 정의되었으며, 부호화 효율을 정량적으로 평가하기 위해서 Bjøntegaard Delta Rate (BD-Rate)<sup>[20]</sup>를 평가 지표로 사용하였다. BD-Rate는 동일한 화질을 기준으로 비트레이트를 비교하여 값이 낮을수록 부호화

효율이 높음을 의미한다.

전체 데이터세트에 대해 기존 DCVC와 제안 방법을 비교하여, 그림 5는 2차원 속성 영상 부호화 성능을 나타내며, 그림 6은 복원된 3차원 포인트 클라우드의 부호화 성능을 나타낸다.

그림 5에서는 입력된 2차원 속성 영상과 2차원 복원 속성 영상 사이의 왜곡을 측정하였으며, 그림 6은 입력된 3차원 포인트 클라우드와 복원된 3차원 포인트 클라우드 사이의 왜곡을 측정하였다. 실험 결과 제안 방법을 DCVC에 적

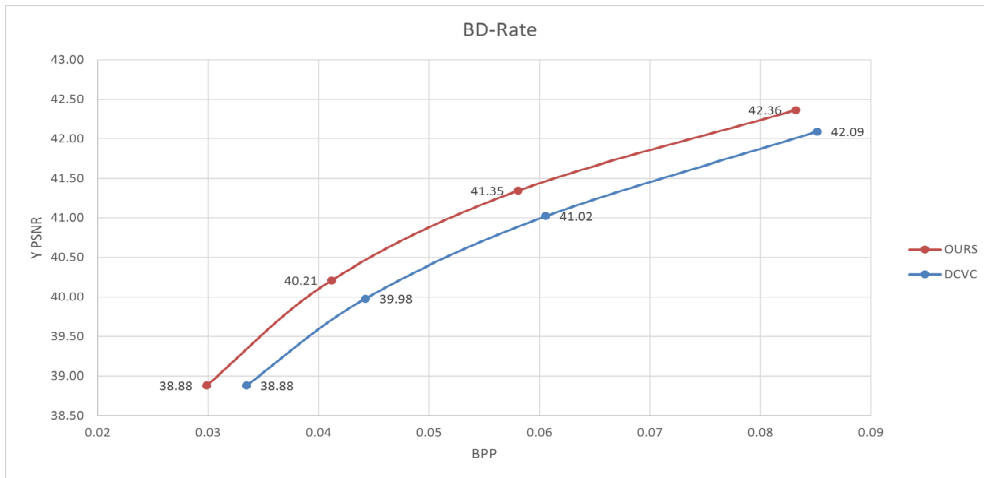


그림 5. 2차원 속성 영상에 대한 휘도 성분의 부호화 효율  
 Fig. 5. Luma coding efficiency for 2D attribute images

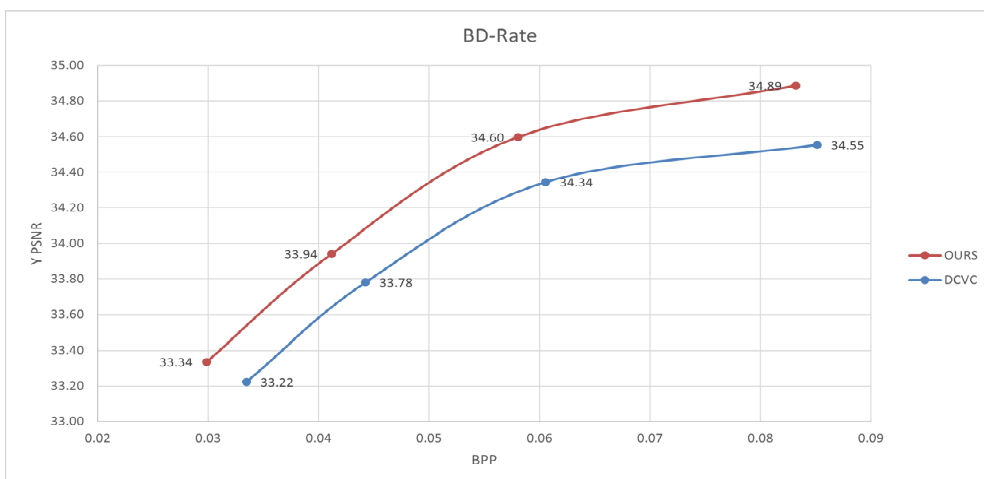


그림 6. 복원된 3차원 포인트 클라우드에 대한 휘도 성분의 부호화 효율  
 Fig. 6. Luma coding efficiency of the reconstructed 3D point cloud

용하였을 때, 2차원 속성 영상의 휘도 성분에 대해 12.6% BD-rate 이득을 보였으며, 복원된 3차원 포인트 클라우드의 휘도 성분에 대해서는 16.7% BD-rate 이득을 보였다. 이는 제안 방법이 비유효 화소에 대한 비트 낭비를 최소화하고 유효 화소에 대한 부호화 효율을 개선하였음을 의미한다. 특히 2차원 속성 영상보다 3차원 복원 품질에서 더 큰 성능 향상이 나타난 것은 제안 방법이 점유맵을 활용하여 실제

3차원 점의 속성 정보를 더 정확하게 복원함을 입증한다. 따라서 제안 방법은 점유 정보를 효과적으로 활용하여 3차원 포인트 클라우드 부호화 성능 개선에 기여할 수 있다.

## 2. Ablation study

제안 방법은 점유맵을 조건부 부호화 모듈의 입력 단계

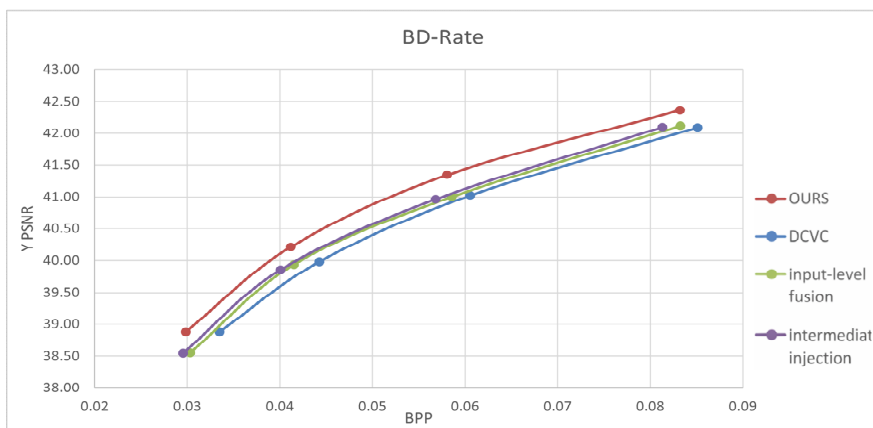


그림 7. 입력 단계에서 속성 영상과 점유맵을 결합하는 방법, 중간 단계에서 점유맵 피처를 주입하는 방법, 그리고 두 방법을 모두 적용한 제안 방법에 대한 2차원 속성 영상 휘도 성분의 부호화 효율 비교

Fig. 7. Coding efficiency comparison for the luma component of 2D attribute images for input-level fusion of attribute images with the occupancy map, intermediate injection of occupancy-map features, and the proposed method combining both schemes

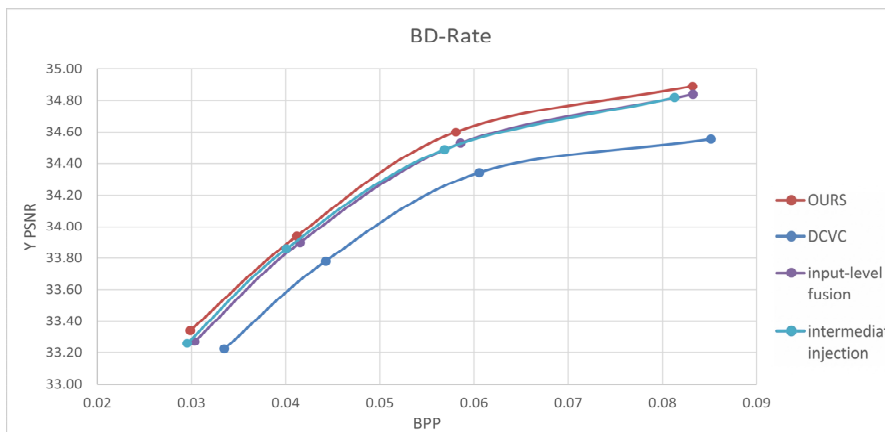


그림 8. 입력 단계에서 속성 영상과 점유맵을 결합하는 방법, 중간 단계에서 점유맵 피처를 주입하는 방법, 그리고 두 방법을 모두 적용한 제안 방법에 대한 복원된 3차원 포인트 클라우드 휘도 성분의 부호화 효율 비교

Fig. 8. Coding efficiency comparison for the luma component of reconstructed 3D point clouds for input-level fusion of attribute images with the occupancy map, intermediate injection of occupancy-map features, and the proposed method combining both schemes

와 중간 단계에서 주입한다. 각 단계의 점유맵 주입이 부호화 성능에 얼마나 영향을 주는지를 분석하기 위해, 입력 단계에서 속성 영상과 점유맵을 결합하는 방법과 중간 단계에서 풀링 기반의 점유맵 피처를 주입하는 방법을 개별적으로 실험하였다.

그림 7은 입력 단계에서 속성 영상과 점유맵을 결합하는 방법, 중간 단계에서 풀링 기반의 점유맵 피처를 주입하는 방법, 제안 방법에 대한 2차원 속성 영상의 휘도 성분에 대한 부호화 성능 결과이다. 입력 단계에서 속성 영상과 점유맵을 결합하는 방법은 3.7% BD-rate 이득을 보였고, 중간 단계에서 풀링 기반의 점유맵 피처를 주입하는 방법은 5.3% BD-Rate 이득을 보였다.

그림 8은 상기 세 가지 방법에 대한 복원된 3차원 포인트 클라우드의 휘도 성분에 대한 부호화 성능 결과이다. 입력 단계에서 속성 영상과 점유맵을 결합하는 방법은 13.4% BD-rate 이득을 보였고, 중간 단계에서 풀링 기반의 점유맵 피처를 주입하는 방법은 14.6% BD-Rate 이득을 보였다. 이는 입력 단계에서는 점유 정보를 명시적으로 활용함으로써 유효 화소의 왜곡을 효과적으로 줄였고, 중간 단계에서는 풀링 기반의 점유맵 피처 주입을 통해 중요한 정보를 강조하여 부호화 효율을 개선하였음을 의미한다. 이는 점유 정보가 조건부 부호화 모듈의 입력 단계나 중간 단계 중 한 위치에서만 적용하더라도 유의미한 성능 향상을 제공할 수 있음을 보여준다.

## V. 결론

기존 V-PCC는 속성 영상과 점유맵이 높은 상관관계를 가지고 있음에도 불구하고, 독립된 스트림으로 분리하여 부호화한다. 이는 속성 영상의 실제 유효 화소 영역을 고려하지 못하기 때문에, 최적의 부호화 성능을 달성하기 어렵다. 본 논문에서는 V-PCC의 속성 영상 부호화 성능을 개선하기 위해 점유맵을 속성 영상 부호화 과정에 주입함으로써 속성 영상과 점유맵 간의 상관관계를 활용하는 방법을 제안하였다. 제안 방법은 점유맵을 조건부 부호화 모듈의 입력 단계와 중간 단계에서 활용한다. 입력 단계에서는 점유맵을 속성 영상과 결합하여 입력으로 사용하고, 중간 단

계에서는 풀링 기반의 블록을 통해 중요한 화소를 강조함으로써 유의미한 유효 화소 정보를 더 잘 학습시킨다. 이를 통해 제안 방법은 속성 영상의 유효 화소에 대한 왜곡을 최소화하도록 유도하고, 점유맵을 활용한 손실 함수를 설계하여 비유효 화소에 대한 중요도를 낮춤으로써 불필요한 비트 낭비를 줄일 수 있다. 실험 결과 제안 방법은 기존 DCVC 대비 2차원 속성 영상의 휘도 성분에 대해 12.6% BD-Rate 이득을 보였으며, 복원된 3차원 포인트 클라우드의 휘도 성분에 대해서는 16.7% BD-Rate 이득을 보였다. 이는 제안 방법이 기존 V-PCC 속성 영상의 부호화 효율을 개선할 수 있음을 의미한다.

향후에는 제안 방법을 속성 영상뿐만 아니라 기하 영상 부호화 과정으로 확장하여 점유맵 정보의 활용 가능성을 분석할 예정이다. 또한 제안 방법을 기반으로 신경망에서의 다양한 점유맵 주입 방법을 연구하고, 이를 DCVC 이외의 다양한 신경망 기반 비디오 부호화 모델에 적용하여 확장성과 범용성을 검증할 예정이다.

## 참고 문헌 (References)

- [1] MPEG, "V-PCC Codec Description," N00100, May 2021.
- [2] MPEG, "Enhanced G-PCC Codec Description," N01234, July 2025.
- [3] J. H. Byun, H. S. Choi, and D. G. Sim, "MPEG G-PCC International Standard Technology," Broadcast and Media Magazine, Vol. 26, No. 2, pp. 31 - 45, 2021 April.
- [4] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the High Efficiency Video Coding (HEVC) Standard," IEEE Transactions on Circuits and Systems for Video Technology, Vol. 22, No. 12, pp. 1649 - 1668, December 2012.  
doi: <https://doi.org/10.1109/TCSVT.2012.2221191>
- [5] J. H. Li, B. Li, and Y. Lu, "Deep Contextual Video Compression," Advances in Neural Information Processing Systems, Vol. 34, pp. 18114 - 18125, December 2021.  
doi: <https://doi.org/10.48550/arXiv.2109.15047>
- [6] B. Bross et al., "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, October 2021.  
doi: <https://doi.org/10.1109/TCSVT.2021.3101953>
- [7] E. S. Jang, "Introduction to MPEG Video-Based Point Cloud Compression Standard," Broadcast and Media Magazine, Vol. 26, No. 2, pp. 18 - 30, 2021 April.
- [8] J. H. Li, B. Li, and Y. Lu, "Neural Video Compression with Diverse Contexts," Proceedings of the IEEE/CVF Conference on Computer

- Vision and Pattern Recognition, pp. 22616 - 22626, June 2023.  
doi: <https://doi.org/10.1109/CVPR52729.2023.02166>
- [9] J. H. Li, B. Li, and Y. Lu, "Neural Video Compression with Feature Modulation," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 26099 - 26108, June 2024.  
doi: <https://doi.org/10.1109/CVPR52733.2024.02466>
- [10] Z. Jia et al., "Towards Practical Real-Time Neural Video Compression," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 12543 - 12552, June 2025.  
doi: <https://doi.org/10.1109/CVPR52734.2025.01170>
- [11] Y. Shi et al., "AlphaVC: High-Performance and Efficient Learned Video Compression," Proceedings of the European Conference on Computer Vision, pp. 616 - 631, October 2022.  
doi: [https://doi.org/10.1007/978-3-031-19800-7\\_36](https://doi.org/10.1007/978-3-031-19800-7_36)
- [12] X. Sheng et al., "NVC-1B: A Large Neural Video Coding Model," arXiv preprint arXiv:2407.19402, July 2024.  
doi: <https://doi.org/10.48550/arXiv.2407.19402>
- [13] B. Liu, Y. Chen, R. C. Machineni, S. Liu, and H. S. Kim, "MMVC: Learned Multi-Mode Video Compression with Block-Based Prediction Mode Selection and Density-Adaptive Entropy Coding," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 18487 - 18496, June 2023.  
doi: <https://doi.org/10.1109/CVPR52729.2023.01773>
- [14] J. Ballé, V. Laparra, and E. P. Simoncelli, "Density Modeling of Images Using a Generalized Normalization Transformation," Proceedings of the International Conference on Learning Representations, April 2016.  
doi: <https://doi.org/10.48550/arXiv.1511.06281>
- [15] MPEG, "V-PCC Test Model v21," N00503, February 2023.
- [16] Y. H. Choi, J. B. Jeong, S. B. Lee, and E. S. Ryu, "Overview of Video-Based Dynamic Mesh Coding (V-DMC) Standard Work," Proceedings of the International Conference on Information and Communication Technology Convergence (ICTC), pp. 578 - 581, Oct 2022.  
doi: <https://doi.org/10.1109/ICTC55196.2022.9952734>
- [17] MPEG, "Common Test Conditions for V-DMC," N00964, July 2024.
- [18] MPEG, "CTC on AI-based point cloud coding, N01122," February 2025.
- [19] MPEG, "Common Test Conditions for V-PCC," N00038, November 2020.
- [20] G. Bjøntegaard, "Calculation of Average PSNR Differences Between RD-Curves," ITU-T SG16, Doc. VCEG-M33, April 2001.

---

## 저 자 소 개



### 이 가 은

- 2024년 : 국립한밭대학교 정보통신공학 (학사)
- 2026년 : 국립한밭대학교 지능미디어공학 (석사)
- ORCID : <https://orcid.org/0009-0008-2142-7034>
- 주관심분야 : 비디오 부호화, 컴퓨터 비전, 딥러닝, 3D 포인트 클라우드, 포인트 클라우드 부호화



### 이 소 은

- 2024년 : 국립한밭대학교 정보통신공학 (학사)
- 2026년 : 국립한밭대학교 지능미디어공학 (석사)
- 2026년 ~ 현재 : 국립한밭대학교 지능미디어공학 박사과정
- ORCID : <https://orcid.org/0009-0000-1580-4886>
- 주관심분야 : 비디오 부호화, 컴퓨터 비전, 비디오 표준화

---

저 자 소 개



**최 해 철**

- 1997년 : 경북대학교 전자공학 (학사)
- 1999년 : 한국과학기술원 전기및전자공학 (석사)
- 2004년 : 한국과학기술원 전기및전자공학 (박사)
- 2004년 ~ 2010년 : 한국전자통신연구원 선임연구원
- 2010년 ~ 현재 : 국립한밭대학교 지능미디어공학과 교수
- ORCID : <https://orcid.org/0000-0002-7594-0828>
- 주관심분야 : 비디오 부호화, 컴퓨터 비전, 딥러닝