

다축 평면 투영 기반 3D 객체 탐지를 활용한 ROI 보존형 포인트 클라우드 다운샘플링

임채운, 박현희*
명지대학교 정보통신공학과

dlacodbs0627, hhpark*@mju.ac.kr

3D Object Detection based on Multi-Axis Plane Projection for ROI-Preserving Point Cloud Downsampling

Lim Chae Yun, Park Hyun Hee*
Dept. Information and Communication Engineering, Myongji University

요 약

본 논문은 3D 포인트 클라우드에서 시각적으로 중요한 영역을 보존하는 다운샘플링을 제안한다. 다축 평면 투영을 기반으로 3D region-of-interest(ROI)를 탐지하고 ROI와 non-ROI를 다른 품질로 다운샘플링한다. 동일 포인트 개수에서 기존 다운샘플링 기법 대비 중요 영역의 세부 구조를 더욱 효과적으로 보존한다. 실험을 통해 다운샘플링 결과의 시각적 품질뿐만 아니라, 탐지 작업을 위한 입력 데이터로도 활용이 가능하다는 것을 입증한다. ROI 보존형 다운샘플링 기법은 적응형 스트리밍에 활용할 수 있고, 핵심 정보를 중심으로 데이터를 전송하는 시맨틱 통신에도 기여할 수 있다.

I. 서 론

포인트 클라우드는 3 차원 공간에 정의된 점들의 집합이다. 객체와 공간을 입체감 있게 시각화할 수 있고, extended reality(XR, 확장현실)로의 연계 가능성 덕에 차세대 미디어 포맷으로 주목받고 있다. 하지만 데이터 크기가 방대하여 무선 통신 대역폭을 초과하는 경우가 발생한다[1]. 데이터 용량 축소 방법으로 포인트 수를 줄이는 다운샘플링 기법이 널리 사용되지만, 핵심 정보까지 함께 제거되는 한계점이 따른다. 특히 가상 통화와 컴퓨터 비전 등 시각적 품질이 중요한 분야에서는 사용자 경험 저하의 원인이 되기 쉽다. 따라서 핵심 정보 손실을 최소화 하면서 포인트의 수를 줄이는 접근법이 요구된다.

본 연구는 다축 평면 투영 기반의 ROI 보존형 다운샘플링 기법을 제안한다. ROI 내의 포인트는 보존하고, non-ROI 내의 포인트는 더 많이 제거해 기존 다운샘플링 기법 대비 시각적으로 중요한 영역의 품질 저하를 최소화한다. 제안하는 기법은 다운샘플링을 활용한 데이터 크기 감소, 핵심 정보 보존이라는 두 목표를 함께 만족한다. 또한 낮은 품질에서도 ROI 정보가 크게 손실되지 않아 시각화 외에 이미지 탐지 작업에서도 안정성을 보인다.



(a) voxel grid downsampling (b) farthest point sampling
(c) ROI-preserving downsampling

그림 1. 기존 기법과 제안 기법의 시각화 비교

II. 본론

그림 1 은 기존 다운샘플링 기법과 제안하는 기법을 동일한 포인트 개수만큼 남긴 시각화 예시이다. 그림 1(a)과 그림 1(b)과 같은 기존 기법들은 전체 포인트를 균일하게 다운샘플링하지만, 얼굴의 상세 정보가 누락되는 문제를 보인다. 반면 제안 기법인 그림 1(c)은 덜 중요한 영역인 몸쪽의 포인트는 더 적은 수로 표현되지만, 중요한 영역인 얼굴의 상세 정보는 더욱 잘 보존한다.

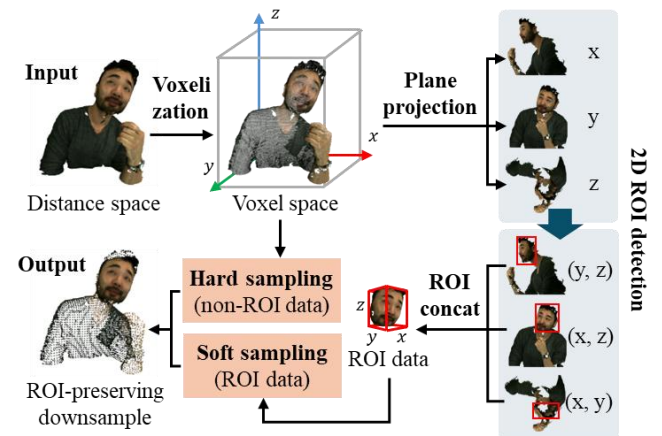


그림 2. ROI 보존형 다운샘플링 기법 전체 아키텍처

그림 2에서의 원본 데이터는 프레임 당 약 150 만 개의 포인트로 인해 22.5MB 라는 매우 큰 용량을 갖는다. 해당 데이터를 미디어 콘텐츠에서 자주 사용하는 30 frames per second(FPS)로 수신하려면 5.4Gbps 의 전송률이 필요하다. 4K 평면 영상의 전송률[1]인 25Mbps 와 비교하면 매우 높은 수치이다. 이는 포인트 클라우드 전송에 데이터 용량 축소가 필수적이라는 것을 보여준다.

본 논문에서 포인트 클라우드의 3D ROI를 얻어내는 방법은 크게 세 단계로 이루어진다. 먼저, 포인트 클라우드를 복셀화하여 균일한 밀도로 변환한다. 원본 포인트 클라우드 데이터는 객체의 특징에 따라 일부 영역에서 과밀집된다. 탐지, 분할 작업에서는 물체의 특징 정보로 활용할 수 있지만 시각화 측면에서는 포인트가 겹치거나 중복되는 문제를 보인다. 다음은 여러 축을 기준으로 평면 투영한다. 3D 포인트 클라우드 데이터를 2D 이미지로 변환하고, 객체 탐지와 같은 딥러닝 모델을 활용해 2D ROI를 구한다. 마지막으로 여러 축의 2D ROI 좌표를 3D 좌표로 통합한다. 하나의 축에서만 ROI가 감지되는 경우, 얻지 못한 좌표는 임의의 최댓값으로 대체한다. 얻어낸 3D ROI 좌표는 시각 정보 보존을 위해 사용한다.

III. 실험

품질 비교에 사용한 voxel grid downsampling 및 farthest point sampling은 포인트 수를 균일하게 줄이는 기존 다운샘플링 기법이다. 여러 품질로 각 기법을 적용해 시각적 품질을 정성적으로 비교한다. 추가로, 제안하는 기법은 다운샘플링 이후에도 주요 정보를 보존할 수 있어 추후 탐지, 분할과 같은 작업에도 강건함을 보일 것으로 예상된다. 이를 실험으로 입증하고자 COCO 데이터셋으로 사전 학습된 YOLOv8을 활용해 80개 클래스 중 person 클래스의 탐지 결과를 정량적으로 측정한다.

데이터셋은 Microsoft Voxelized Upper Bodies - david 포인트 클라우드 영상을 활용한다. 해당 영상은 30FPS의 주 사율을 갖는다. 2D ROI 탐지 모델은 WIDAR FACE 얼굴 데이터셋으로 파인튜닝된 YOLOv8을 사용한다. 다운샘플링 품질의 경우 국내 통신 3사 서비스 품질 평가에서의 5G 평균 다운로드 속도인 1,025.52 Mbps[2]를 실험의 최대 품질 기준으로 삼는다. 이 경우 30FPS 조건에 맞추기 위해 한 프레임에 수용할 수 있는 최대 포인트 수는 284,866 포인트이며, 이를 최대 품질(resolution) r_{max} 라 정의한다. 수식 1에서 r_k 는 r_{max} 에 대해 k 단계 강도의 다운샘플링을 적용한 포인트 수를 의미한다. k 가 증가할수록 포인트 수는 지수적으로 감소하며, 소수점 이하는 내림 처리하여 정수로 표현한다.

$$r_k = \left\lfloor \frac{r_{max}}{2^k} \right\rfloor, \quad k = 0, 1, 2, 3, 4 \quad (1)$$

Resolution	$r_0(r_{max})$	r_1	r_2	r_3	r_4
Voxel grid					
Farthest point Sampling					
ROI preserving (Ours)					
Point num	284,866	142,433	71,216	35,608	17,804
Bitrate @30fps (Mbps)	1,026	513	256	128	64

그림 3. 단계적 다운샘플링 시각화 및 전송률 품질 비교

그림 3은 단계별 시각화 품질과 함께 30FPS로 전송하는 데 필요한 전송률을 보여준다. 다운샘플링 기법의 하이퍼파라미터는 각 기법에 맞게 적절하게 설정한다. voxel grid downsampling 및 farthest point sampling 기법은 품질 단

계가 낮아질수록 시각 정보가 전체적으로 소실되어 중요 영역의 세부 정보는 식별하기 어려워진다. 반면 제안하는 기법은 품질이 낮아지더라도 중요 영역에 대한 시각 정보를 유지한다.

표 1. YOLOv8 모델 기반 person 클래스 탐지 예측 결과

Type	Evaluation	Resolution					
		r_0	r_1	r_2	r_3	r_4	avg
Voxel grid	POD	1.0	1.0	0.833	0.097	0.222	0.631
	Confidence	0.903	0.864	0.649	0.322	0.398	0.627
Farthest point sampling	POD	1.0	0.963	0.458	0.051	0.153	0.525
	Confidence	0.892	0.866	0.675	0.363	0.434	0.646
ROI preserving (Ours)	POD	1.0	1.0	0.671	0.727	0.736	0.827
	Confidence	0.905	0.864	0.637	0.543	0.517	0.693

표 1은 각 기법으로 다운샘플링된 결과 이미지를 탐지하여 평균 낸 결과이다. probability of detection(POD)은 person 객체가 성공적으로 감지된 확률이다. confidence는 감지된 person 객체의 신뢰도 점수이다. 기존 기법들은 데이터 품질에 따라 POD과 confidence가 크게 불안정하다. 반면 본 논문에서 제안하는 기법은 전체적인 품질에서 탐지 작업을 안정적으로 수행한다.

IV. 결론

본 논문은 다축 평면 투영으로 얻은 3D ROI를 활용해 중요 영역을 보존하는 포인트 클라우드 다운샘플링을 제안한다. 동일 포인트 개수에서 기존 기법 대비 시각적으로 중요한 영역을 효과적으로 보존한다. 또한 다양한 품질로 얻어낸 이미지를 활용해 탐지 작업을 수행했을 때 POD와 confidence에서 안정적인 값을 보여준다.

ROI 보존형 다운샘플링은 별도 코덱 없이도 원본 데이터의 용량을 효과적으로 감축할 수 있다. 다양한 품질에서도 중요 정보를 효과적으로 보존하는 점에서 적응형 스트리밍 또는 시맨틱 통신에 활용할 수 있을 것으로 보인다. 다만 현재로서는 ROI 탐지 모델 성능에 따라 결과가 크게 변화하고, 시간 흐름 정보를 반영하지 않아 프레임 간 일관성이 낮을 수 있다는 한계점이 따른다.

향후 연구에서는 다중 클래스 탐지 또는 문맥 파악을 활용해 중요도 기반 다운샘플링으로 확장하고자 한다. 또한 사람 데이터 외에도 LiDAR 센서 기반 도시 데이터와 같이 대규모 포인트 클라우드에서의 다축 평면 투영 및 ROI 보존형 다운샘플링을 적용해 볼 계획이다.

ACKNOWLEDGMENT

이 논문은 산업통상자원부 및 산업기술평가원(KEIT)과 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임 (No. RS-2024-00469138, 국산 AI 반도체 기반 비전인식기술 소프트웨어개발도구 개발, No. 2022R1A2C2005705, 분산 머신러닝 기반 지능형 플라잉 기지국을 위한 AI-MAC 프로토콜)

참 고 문 헌

- [1] Z. Liu et al., "Point Cloud Video Streaming: Challenges and Solutions," in IEEE Network, vol. 35, no. 5, pp. 202-209, Sept. 2021, doi: 10.1109/MNET.101.2000364
- [2] 과학기술정보통신부·한국지능정보사회진흥원. "2024년 통신서비스 수신 가능 지역 점검 및 품질평가 결과 발표." 보도자료. Dec, 2024.