치수 측정을 위한 3D 복원의 과업 지향 재설계: 상대 기하 보존 기반 효율화

김이든°[1], 안준식[1], 김우주[2], 이다영[1*]
강원대학교 [1]컴퓨터공학과, [2]데이터사이언스학과
{rladlems1031, fcsolution11, woojoo.kim, dayoung08}@kangwon.ac.kr

Task-Oriented Redesign of 3D Reconstruction for Dimensional Measurement: Efficiency through Relative Geometry Preservation

Yideun Kim°[1], Joonsik An[1], Woojoo Kim[2], Dayoung Lee[1*]
[1]Department of Computer Science and Engineering, [2]Graduate School of Data Science
Kangwon National University

요 약

단일 이미지 기반 3차원 복원은 범용 시각 품질 향상을 중심으로 발전해 왔지만, 치수 측정 응용에서는 높은 계산 비용과 메모리 사용량이 부담이 될 수 있다. 본 논문은 SAM 3D를 치수 측정 과업에 맞게 재구성하여, 외곽 형상과 상대 기하 구조를 유지하면서 불필요한 연산을 줄인 경량화 파이프라인을 제안한다. 이를 위해 고정 깊이 기반 포인트 맵, 구조 생성 단계의 계산 재사용, 추론 반복 수 축소, 메시 복원 제거, 가우시안 기반 직접 치수 산출을 적용하였다. 실험 결과, 제안 방법은 치수 측정에 필요한 정확도를 유지하면서도 원본 SAM 3D 대비 추론 시간과 VRAM 사용량을 크게 줄였고, Fast-SAM 3D와 비교해서도 추가적인 자원 효율 개선을 보였다. 이는 치수 측정에서는 범용 복원 품질의 극대화보다 상대 기하 보존 중심의 과업 지향 재설계가 효과적일 가능성을 시사한다.

1. 서론

이사 물류, 실내 공간 계획, 중고 거래, 가구 배치와 같은 응용에서는 사진 한 장으로 객체의 상대적 치수를 빠르게 추정할 수 있는 기술이 중요하다. 최근 단일 이미지 기반 3차원 복원은 새로운 시점 생성 기반 접근에서 출발해 형상, 질감, 배치를 함께 복원하는 방향으로 발전해 왔으며, SAM 3D는 복잡한 자연 이미지에서도 강건한 복원 성능을 보이며 그 적용 가능성을 크게 넓혔다[1, 2]. 그러나 기존 연구는 주로 시각적 품질과 메시 완성도 향상에 초점을 두어, 치수 측정에 필요한 기하 정보를 짧은 시간과 제한된 자원 안에서 제공하는 문제는 충분히 다루지 못했다.

Fast-SAM 3D는 SAM 3D의 계산량을 줄여 추론 속도를 개선하였으나, 여전히 범용 3차원 복원을 더 빠르게 수행하는 데 초점을 둔다[3]. 반면 치수 측정에서는 고품질 복원 자체보다, 외곽 형상과 상대적 기하 구조를 유지하면서 불필요한 표현과 연산을 제거하는 것이 더 직접적인 목표가 된다. 또한 PMODE와 같은 단일 이미지 기반 치수 추정은 3차원 형상을 명시적으로 복원하지 않아 복잡한 외곽 구조를 다루는 데 한계가 있다[4]. 따라서 치수 측정을 위해 3차원 복원 파이프라인 자체를 목적에 맞게 재구성하는 접근이 필요하다.

이에 본 연구는 병목 요소를 제거하거나 단순화하여, 치수 측정에 필요한 기하 정보만 유지하는 3차원 복원 구조를 제안한다. 그 결과, 정확도는 유사하게 유지하면서 추론 시간과 VRAM 사용량을 줄였고, 치수 측정에서는 범용 복원 품질보다 상대 기하 보존 중심의 재설계가 더 효과적일 가능성을 보였다.

*) 본 논문은 2026년도 과학기술정보통신부 및 정보통신기획평가원의 'SW중심대학사업'(2025-0-00058)의 지원을 받아 수행된 연구결과입니다.

2. 제안 방법

본 연구는 SAM 3D의 추론 파이프라인을 치수 측정 목적에 맞게 재구성한다[2]. 원본 SAM 3D는 범용 3차원 복원에는 효과적이지만, 치수 측정에서는 시각적 품질 향상을 위한 연산까지 포함하므로 추론 시간과 메모리 사용량이 증가할 수 있다.

본 연구의 목표는 입력 이미지와 객체 마스크로부터 치수 추정에 필요한 기하 정보를 안정적으로 확보하고, 객체 치수 $\hat{\mathbf{d}}$를 산출하는 것이다. 이를 위해 입력 준비 단계의 깊이 예측, 반복 추론, GPU 메모리 점유, 출력 디코딩을 주요 병목으로 식별하고, 치수 측정에 직접 필요하지 않은 연산을 제거한다.

2.1 고정 깊이 기반 점 지도

원본 SAM 3D는 MoGe 기반 깊이 추정을 통해 3차원 포인트 맵을 생성한다[5]. 그러나 본 연구는 절대 깊이 복원이 아니라 상대 치수 비율과 외곽 형상 보존을 목표로 하므로, 별도의 깊이 추정 대신 고정 깊이와 핀홀 카메라 모델에 기반한 포인트 맵을 사용한다[6].

카메라 내부 파라미터 행렬 K는 초점거리 f와 주점 (c_x, c_y)로 구성된다. 이미지 크기 $H \times W$에 대해 주점은 $c_x = (W-1)/2$, $c_y = (H-1)/2$로 두고, 초점거리 f는 이미지 크기에 비례하도록 설정한다. 모든 픽셀에 동일한 깊이값 z를 적용하면, 픽셀 좌표 (u, v)의 3차원 점 $\mathbf{p}(u, v)$는 다음과 같이 정의된다.

$$\mathbf{p}(u,v) = zK^{-1}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{u-c_x}{f}z \\ \frac{v-c_y}{f}z \\ z \end{bmatrix} \quad (1)$$

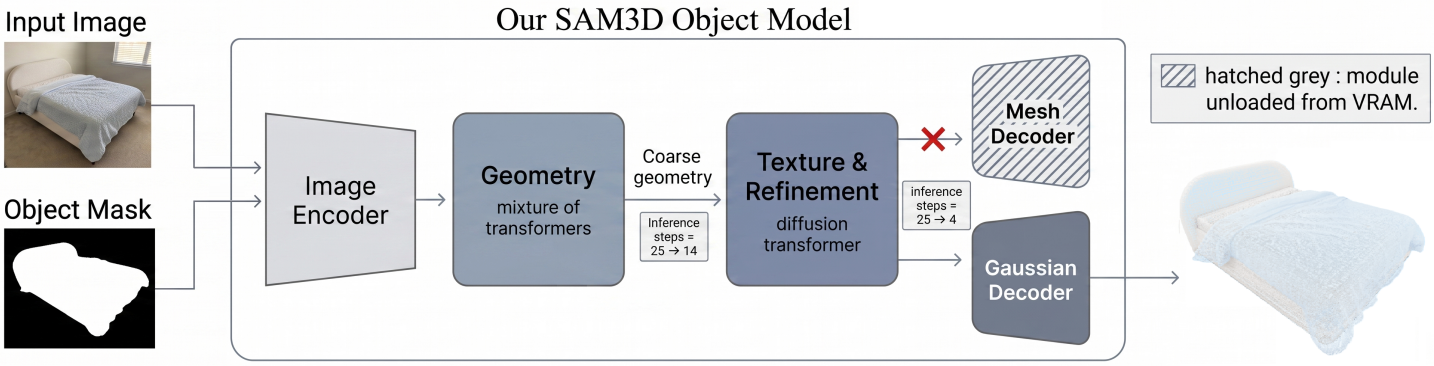


그림 1: 치수 측정 목적에 맞게 상대 기하 보존 중심으로 재구성한 SAM 3D 추론 파이프라인

이 방식은 절대 깊이는 제공하지 않지만 상대적 기하 구조를 구성하는 데 충분하며, 깊이 추정 추론을 제거함으로써 입력 단계의 계산량과 메모리 사용량을 줄인다.

2.2 구조 생성 단계의 계산 재사용

구조 생성 단계는 SAM 3D의 반복적 복원 과정에서 전체 추론 시간의 큰 비중을 차지한다[2]. 본 연구는 Fast-SAM 3D에서 제안된 단계 캐싱 기법을 구조 생성 단계에 한정하여 적용한다[3]. 재계산이 수행되는 시간 단계의 집합을 $\mathcal{C}$라 하고, 시간 단계 t에서 입력 상태 x_t에 대해 계산된 백본 출력을 $f_\theta(x_t, t)$라 하면, 실제 추론에 사용되는 출력 v_t는 다음과 같이 정의된다.

$$v_t = \begin{cases} f_\theta(x_t, t), & t \in \mathcal{C} \\ v_{t-1}, & \text{otherwise} \end{cases} \tag{2}$$

이는 선택된 단계에서만 출력을 새로 계산하고, 나머지 단계에서는 이전 단계의 출력을 재사용함을 의미한다. 본 연구에서는 초기 두 단계는 모두 계산한 뒤, 이후에는 세 단계 간격으로만 출력을 재계산한다.

2.3 반복 수 및 메모리 점유 축소

치수 측정에서는 세밀한 질감보다 외곽 형상과 상대 크기 추정이 더 중요하므로, 구조 생성 단계와 세밀화 단계의 추론 step을 각각 $25 \to 14$, $25 \to 4$로 축소하였다. 구조 생성 단계는 기하 정보 보존을 위해 완만하게, 세밀화 단계는 영향이 제한적이므로 더 크게 축소하였다. 또한 치수 산출에 직접 사용되지 않는 메시 복원 모듈과 관련 서브 모듈을 제거하여 메모리 점유와 후처리 비용을 함께 줄였다.

2.4 가우시안 기반 직접 치수 산출

본 연구는 메시 디코딩을 생략하고, 가우시안 좌표로 구성한 포인트 집합으로부터 OBB (Oriented Bounding Box)를 계산하여 객체 치수를 산출한다. 가우시안 중심점 집합을 $\mathcal{G} = \{\mathbf{g}_i \in \mathbb{R}^3\}_{i=1}^N$라 하고, OBB의 축 길이를 $\mathbf{e} = [e_1, e_2, e_3]^\top$, 회전행렬을 $R \in \mathbb{R}^{3\times3}$라 하자. 각 OBB 축과 좌표축의 정렬 정도를 기준으로, 각 OBB 축을 가장 잘 정렬된 좌표축에 일대일로 대응시키는 함수 π를 구성하면, 객체 치수 $\hat{\mathbf{d}}$는 다음과 같이 계산된다.

$$\hat{\mathbf{d}} = \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix}, \quad d_x = e_{\pi(x)},\ d_y = e_{\pi(y)},\ d_z = e_{\pi(z)} \tag{3}$$

이 방식은 메시 복원 없이도 치수 추정에 필요한 기하 정보를 직접 활용할 수 있어 전체 추론 시간을 줄인다.

3. 실험

3.1 실험 환경

본 연구는 SAM 3D의 두 모델 중 `sam-3d-objects`를 최적화 대상으로 선정하였다. 데이터셋은 고해상도 2D 이미지와 이에 대응하는 3D 모델, 이미지, 객체 마스크를 제공하는 Pix3D [7]를 사용하였다. 전체 9개 가구 카테고리, 10,069개 샘플 중 잘림 또는 가림이 존재하거나 마스크 면적이 1,000 픽셀 미만인 샘플을 제외하여 7,021개를 확보하였고, 이후 각 카테고리의 비율을 유지하여 최종 500개를 실험 대상으로 선정하였다. 모든 실험은 NVIDIA L4 GPU (24GB VRAM) 환경에서 수행하였다.

3.2 평가 지표

본 연구는 치수 측정 성능과 자원 효율성을 함께 평가하기 위해 정확도 지표와 효율성 지표를 모두 사용하였다. 치수 측정 성능은 Relative Dimension Error (RDE), Dimension Accuracy@5 (DA@5), Volume Error (VE)를 중심으로 평가하였고, 전체 형상 유사도는 보조 지표로 Chamfer Distance (CD)를 함께 측정하였다. 자원 효율성은 평균 VRAM 점유량과 평균 추론 시간을 기준으로 평가하였다.

정확도 지표는 예측 객체와 Ground Truth 3D OBB를 정렬한 뒤

계산하였다. 이는 제안 기법이 OBB 기반 치수 산출을 수행하므로, 예측 치수와 정답 치수를 동일한 기하 기준에서 비교하기 위함이다. 치수는 크기순으로 재정렬하여 축 대응의 모호성을 줄였으며, CD는 각 형상의 최장 축을 기준으로 $[-0.5, 0.5]$ 범위로 정규화한 뒤 측정하였다.

- **RDE (%):** 예측 치수와 실제 치수 간의 상대 오차
- **DA@5 (%):** 예측 치수가 실제 치수의 $\pm 5\%$ 범위 내에 포함되는 비율
- **VE (%):** 예측 객체의 부피와 실제 부피 간의 상대 오차
- **CD ($\times 10^{-3}$):** 예측 3D 형상과 실제 3D 형상 간 양방향 최근접점 거리의 대칭 평균
- **평균 VRAM 점유량 (GB):** 객체당 추론 과정에서 사용된 최대 GPU 메모리 용량의 평균
- **평균 추론 시간 (s):** 객체 하나를 처리하는 데 소요된 평균 추론 시간

3.3 실험 결과 및 분석

표 1와 표 2는 정확도와 자원 효율성 비교 결과를 나타낸다.

표 1: 치수 측정 성능 비교

	RDE↓	DA@5↑	VE↓	CD↓
Ours	**8.588**	**33.400**	**24.523**	**0.002195**
SAM 3D	8.277	35.500	22.971	0.001893
Fast-SAM 3D	8.474	32.400	23.393	0.002348

표 2: 자원 효율성 비교

	평균 VRAM 점유량↓	평균 추론 시간↓
Ours	**11.14 GB**	**5.57 s**
SAM 3D	17.62 GB	55.27 s
Fast-SAM 3D	17.22 GB	7.23 s

3.4 실험 결과 및 분석

표 1와 표 2는 원본 SAM 3D, Fast-SAM 3D, 그리고 제안 기법의 치수 측정 성능과 자원 효율성을 비교한 결과를 보여준다. 제안 기법은 원본 SAM 3D 대비 평균 추론 시간을 약 89.9%, VRAM 사용량을 약 36.8% 감소시켰으며, Fast-SAM 3D와 비교해서도 평균 추론 시간은 약 23.0%, VRAM 사용량은 약 35.3% 추가로 감소시켰다. 이는 제안 기법이 자원 효율성 측면에서 가장 큰 이점을 가짐을 보여준다.

원본 SAM 3D는 모든 정확도 지표에서 가장 우수하여, 정확도만 기준으로 보면 가장 강한 기준선이다. 반면 제안 기법은 원본 대비 일부 정확도 저하를 보였는데, 이는 범용 복원 품질의 극대화보다 치수 측정에 필요한 상대 기하 정보의 효율적 확보를 목표로 파이프라인을 재구성하였기 때문이다.

기존 경량화 기준선인 Fast-SAM 3D와 비교하면, 제안 기법은 RDE와 VE에서는 다소 불리하였으나, DA@5는 소폭 증가하였고 CD는 약 6.5% 감소하여 더 좋은 성능을 보였다. 이는 평균 오차는 일부 증가하더라도, 허용 오차 범위 내 치수 정확도와 형상 유사도는 개선될 수 있음을 보여준다.

종합하면, 제안 기법은 원본 SAM 3D와 Fast-SAM 3D 대비 더 낮은 추론 시간과 VRAM 사용량을 달성하면서도, 치수 측정에 활용 가능한 수준의 품질을 보였다.

4. 결론

본 연구는 SAM 3D의 추론 파이프라인을 치수 측정 목적에 맞게 재구성하여, 측정에 직접 필요하지 않은 연산을 제거하고 상대 기하 보존 중심의 경량화 구조를 제안하였다. 실험 결과, 제안 방법은 치수 정확도를 실용 가능한 수준으로 유지하면서도 원본 SAM 3D 대비 추론 시간과 VRAM 사용량을 크게 줄였으며, Fast-SAM 3D와 비교해서도 비슷한 수준의 정확도를 유지하면서 추가적인 자원 효율 개선을 보였다. 이는 치수 측정 응용에서 범용 3차원 복원 품질의 극대화보다, 외곽 형상과 상대 기하 구조를 보존하는 범위 내에서 불필요한 연산을 제거하는 과업 지향적 재설계가 더 효과적일 수 있음을 시사한다.

다만, 본 연구의 결과만으로 제안 방법이 실제 치수 측정 응용에 충분한 정확도를 제공한다고 단정하기는 어렵다. 본 실험은 주로 기존 방법 대비 상대적인 정확도와 자원 효율의 변화를 비교하는 데 초점을 두었으며, 실제 활용을 위한 절대적 허용 오차 기준은 별도로 검증하지 못하였다. 따라서 제안 방법의 실질적 유효성은 응용 환경별 요구 정확도를 반영한 추가 평가를 통해 보완될 필요가 있다. 향후 연구에서는 실제 사용 시나리오를 반영한 허용 오차 기준을 설정하고, 정확도와 자원 효율 간의 균형 관계를 응용 목적별로 정량적으로 분석할 예정이다.

참고 문헌

[1] R. Liu *et al.*, "Zero-1-to-3: Zero-shot One Image to 3D Object," in *Proc. IEEE/CVF ICCV*, pp. 4397–4407, 2023.

[2] SAM 3D Team *et al.*, "SAM 3D: 3Dfy Anything in Images," *arXiv preprint arXiv:2511.16624*, 2025.

[3] W. Feng *et al.*, "Fast-SAM3D: 3Dfy Anything in Images but Faster," *arXiv preprint arXiv:2602.05293*, 2026.

[4] T. Khalid *et al.*, "PMODE: Prototypical Mask based Object Dimension Estimation," *arXiv preprint arXiv:2212.13281*, 2022.

[5] R. Wang *et al.*, "MoGe: Unlocking Accurate Monocular Geometry Estimation for Open-Domain Images with Optimal Training Supervision," in *Proc. IEEE/CVF CVPR*, pp. 5261–5271, 2025.

[6] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2 ed., 2004.

[7] X. Sun *et al.*, "Pix3D: Dataset and Methods for Single-Image 3D Shape Modeling," in *Proc. IEEE/CVF CVPR*, pp. 2974–2983, 2018.