

자동 생성 Trimap 기반 매팅&인페인팅 파이프라인을 활용한 배경 전용 3D Gaussian Splatting

이진혁, 황다희, 김동호*

서울과학기술대학교

(jinlee382, haei4061, dongho.kim*)@seoultech.ac.kr

Background-only 3D Gaussian Splatting with an auto-generated Trimap-based matting & inpainting pipeline

Lee Jin-Hyeok, Hwang Da-Hee, Kim Dong-Ho*

Seoul National University of Science & Technology

요약

본 논문에서는 Trimap 기반의 객체 분리와 최적화된 인페인팅(Inpainting) 기법을 결합하여, 2D 이미지에서 원하는 객체를 제거하고 배경만으로 구성된 3D 공간을 효율적으로 재구성하는 새로운 파이프라인을 제안한다. 현재 다양한 공간 미디어 압축/생성 기법이 포인트 클라우드와 입력 이미지를 기반으로 하고, 후처리 알고리즘은 개별 객체에 대하여 진행되는 가운데, 포인트 클라우드 생성 단계에서 객체와 배경을 분리 생성하는 작업이 없어 포인트 클라우드를 기반으로 생성된 공간미디어에서 후처리로 객체와 배경을 분리해야 한다. 이러한 문제를 해결하기 위해 포인트 클라우드 생성단계에서 선행 연구된 Trimap 생성 기법을 활용하여 입력 이미지단에서 객체와 배경을 분리하는 모델을 제안하고, 실험을 통해 실제 동작 가능한 모델임을 보인다.

I. 서론

최근 영상 기반 3차원 장면 재구성 연구에서 전경 객체와 배경을 분리하는 과정은 매우 중요한 전처리 단계로 자리 잡았다. 일반적인 분할(segmentation) 기법은 객체 경계 주변에서 계단 현상이 발생하거나, 미세한 경계 정보를 잃기 쉽지만, 매팅(matting) 기법은 픽셀 단위의 알파(α) 값을 활용해 경계부의 반투명 정보까지 보존한다. 특히 Trimap 기반 매팅 방식은 전경·배경·미분류(unknown) 영역을 명시적으로 정의하여 ([그림 1] 참조), 복잡한 경계부에서도 높은 정확도의 알파 맵을 추론할 수 있다.



그림 1. 입력 이미지와 사용자 지정 Tri-map[3]

현존하는 이미지 매팅 기법은 대체로 추출할 객체에 대해 사용자가 직접 생성한 Trimap[1] 또는 scribble[2]을 필요로 한다. 이로 인해 Trimap 생성 절차를 포함한 매팅 전 과정을 자동화하는 연구[4]가 등장했지만, 해당 방식들은 단일 객체가 명확히 구분되는 이미지에서만 안정적인 결과를 보이며, 다수의 객체가 존재하는 이미지에서는 한계점을 보였다.

따라서 본 연구는 사용자 지시를 반영한 Trimap 생성 방법을 제안하고, 해당 Trimap을 기반으로 한 매팅 과정을 통해 전경을 분리한다. 분리된 전경 영역을 인페인팅(Inpainting) 기법으로 지워 배경만으로 이루어진 이미지를 획득한 뒤, 이 배경 이미지들을 3D Gaussian Splatting에 적용하여 배경 전용 3D 장면을 생성한다.

II. 본론

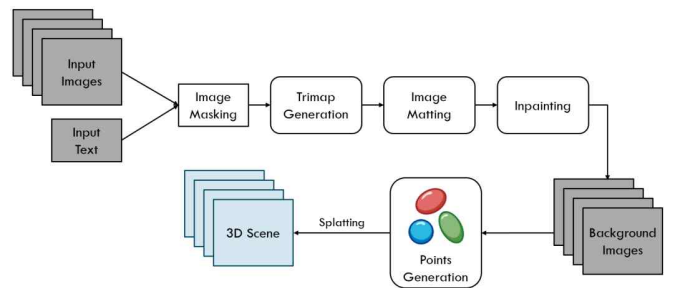


그림 2. 배경 전용 3D 장면 재구성 파이프라인

입력 이미지와 사용자 지시문으로부터 시작해 Trimap 자동 생성, 매팅, 인페인팅 작업을 거쳐 전경이 제거된 이미지를 생성한다. 그 후에 3D Gaussian Splatting을 활용한 3D 배경 장면 표현을 생성하여 배경만 존재하는 novel view synthesis를 제안한다. 제안하는 배경 전용 3D 장면 재구성 파이프라인의 전체 동작도는 [그림 2]와 같다.

A) Trimap 자동 생성

입력 이미지 I 에서 Mask2Former로 얻은 여러 객체 마스크 중 CLIP 기반 분류로 사용자 지시문과 가장 유사한 마스크를 선택한 뒤, 축소 및 확장 연산으로 잡음을 제거해 definite foreground F 와 unknown 영역 U 를 정의한다. 이들로 생성되는 Trimap T 는 두 단계로 나뉘어 계산되며, 알고리즘 동작 과정은 [그림 3]과 같다.

$$T_0(p) = F_0(p) \cdot 255 + (M^+(p) - F_0(p)) \cdot 128 \quad (1)$$

$$T_1(p) = F_1(p) \cdot 255 + (U^1(p) - F_1(p)) \cdot 128 \quad (2)$$

여기서 F_0, F_1 은 각각 축소 및 최종 축소된 definite foreground 마스크를 의미하며, M^+, U_1 은 처음 확장된 unknown 후보 및 경계부 윤곽 정제를 거친 unknown 마스크이다. 마지막으로 unknown 대비 foreground

면적 비율을 계산하고, 사전 설정한 신뢰도 범위 내에 들도록 본 논문에서는 F_1 과 U_1 을 반복 확장 및 축소하여 최종 Trimap을 얻는다.

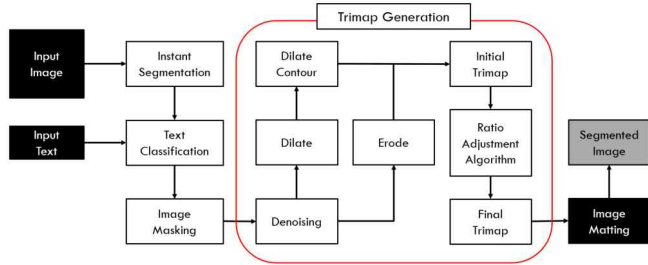


그림 3. Trimap 자동 생성 알고리즘 동작도[5]

B) Inpainting

Trimap의 definite foreground(F_1) 영역을 기준으로 전경 객체를 제거하고 IOPaint 통해 전경 객체가 자연스럽게 지워진 배경 전용 이미지 세트(I_b)를 획득한다.

C) 3D Gaussian Splatting(3DGS)

본 연구에서는 전경 객체가 제거된 이미지 세트(I_b)를 입력으로, 먼저 Structure-from-Motion 알고리즘을 통해 camera pose와 3D point cloud를 정밀하게 추정한다. 이어서 이 points를 중심,공분산,색상 그리고 투명도 파라미터를 지닌 Gaussian으로 변환하고, photometric loss를 최소화하도록 경사하강법 기반 최적화를 수행한다. 최종적으로 최적화된 Gaussian 군을 GPU 친화적 rasterization으로 타원형 splat으로 투영·블렌딩해, 객체가 제거된 배경 전용의 3D 장면 표현을 생성한다[6].

III. 실험 및 결과

본 논문에서는 'NERDS 360 Multi-View dataset for Outdoor Scenes' [7] 데이터셋을 사용하였다. 해당 데이터셋은 각 scene마다 90개의 서로 다른 시점의 이미지를 제공하며, 복수의 객체가 포함되어 있어 매팅 성능 평가에 적합하다. 이를 바탕으로 제안 기법의 매팅 성능 검증 실험 및 배경 이미지 생성 실험을 진행하였다([그림 4] 참조).

[그림 4]은 복잡한 도시 도로 장면을 대상으로 제안된 파이프라인의 중간 및 최종 출력을 보여준다. 왼쪽(a)에서는 여러 대의 차량이 혼재된 원본 이미지를 확인할 수 있고, 중앙(b)에서는 Trimap 자동 생성 및 매팅 과정을 통해 자동차 전경이 정확히 분리된 마스크 이미지를 나타낸다. 오른쪽(c)는 해당 마스크를 기반으로 IOPaint를 적용해 차량이 자연스럽게 사라지고 차선·도로 텍스처가 매끄럽게 보인 최종 배경 이미지이다. 경계부 artifact 없이 주변 배경과 일관된 결과를 얻었으며, 이렇게 얻어진 배경 이미지를 3DGS에 투입하면 [그림 5]와 같이 배경 3차원 장면을 재구성할 수 있다.

하지만 전체 실험과정에서 인페인팅의 의존도가 높고, 완전히 가려진 영역의 배경 정보에 대한 복원이 어려운 것을 확인했다. 이는 Trimap 자동 생성 매팅을 사용하여 이미지 내 객체를 분할 하지만, 그림자, 반사 등 주변 부수 객체까지 완전히 지우지 못하기 때문이다([그림 6] 참조). 이로 인해 3D Gaussian Splatting 단계에서 해당 영역에 artifact가 남아 장면 품질이 저하될 수 있는 한계를 확인했다.

III. 결론

본 연구는 사용자 지시 기반 Trimap 자동 생성 기법을 통해 복잡한 다객체 이미지에서 원하는 객체의 경계를 정확히 분할하고, IOPaint 인페인팅으로 전경을 제거한 배경 이미지를 획득한 뒤, 3D Gaussian Splatting을 적용해 배경 전용 3D 장면을 재구성하는 통합 파이프라인을 제안한다. 이를 통해 기존 분할 기법 대비 경계부 artifact를 억제하고, Trimap 기반

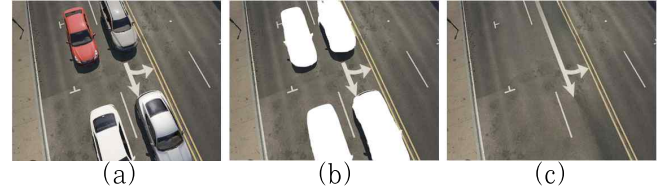


그림 4. (a) Input image, (b) matting image (c) background image



그림 5 3D 배경 장면 재구성 이미지



그림 6. 객체 외의 부수적인 영역 예시

매팅 모델을 다양한 이미지에 적용할 수 있으며, 메타버스·AR/VR 등 3차원 공간 표현 응용에서 배경 재구성 효율성과 렌더링 품질을 동시에 개선할 수 있는 잠재력을 제시하였다.

ACKNOWLEDGMENT

이 논문은 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원·대학ICT연구센터(ITRC)의 지원을 받아 수행된 연구임 (IITP-2025-RS-2021-II211816)

참 고 문 헌

- [1] Ning Xu, Brian Price, Scott Cohen, Thomas Huang; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [2] J. Wang and M. F. Cohen, "An iterative optimization approach for unified image segmentation and matting," Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1, Beijing, China, 2005.
- [3] Christoph Rhemann, Carsten Rother, Jue Wang, Margrit Gelautz, Pushmeet Kohli, Pamela Rott, A Perceptually Motivated Online Benchmark for Image Matting. Conference on Computer Vision and Pattern Recognition (CVPR), June 2009.
- [4] C. -L. Hsieh and M. -S. Lee, "Automatic trimap generation for digital image matting," 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, Kaohsiung, Taiwan, 2013.
- [5] 황다희, 서봉석, and 김동호, "멀티모달 모델과 이미지 분할 모델을 활용한 지시 기반 이미지 매팅을 위한 Trimap 생성 방법," 한국방송미디어공학회 학술발표대회 논문집 (2024): 334-337.
- [6] Kerbl, Bernhard, et al. "3d gaussian splatting for real-time radiance field rendering." ACM Trans. Graph. 42.4 (2023): 139-1.
- [7] Irshad, Muhammad Zubair, et al. "Neo 360: Neural fields for sparse view synthesis of outdoor scenes." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023.