폐색이 고려된 3 차원 재구성을 위한 에이모달 깊이 추정에서 깊이 일관성 향상 기법

이두열, 이채은
인하대학교 전기컴퓨터공학과

dylee910@inha.edu, chae.rhee@inha.ac.kr

Depth Consistency Enhancement in Amodal Depth Estimation for Occlusion-aware 3D Reconstruction

Du Yeol Lee, Chae Eun Rhee

Department of Electrical and Computer Engineering

Inha University

요 약

실사 이미지로부터 3 차원 장면을 재구성하는 과정에 있어 RGB 이미지로부터 카메라의 거리를 나타내는 깊이 맵을 추정하는 것은 중요한 작업이다. 이러한 깊이 맵은 최근 Deep Learning Model 의 발달로 인해 고비용의 장비를 사용하지 않고 출력할 수 있게 되었다. 그러나 전통적인 깊이 추정에서는 물체와 물체간이 서로 가려지는 폐색 환경이 충분히 고려되지 않으며 이는 깊이를 입력 받는 3 차원 재구성 모델의 성능과 3 차원 결과물의 품질을 저해하게 된다. 반복적 깊이 예측 구조를 통해 폐색 영역의 깊이를 출력하여 개선한 기존 연구가 있으나 반복적으로 출력된 깊이맵들의 일관성이 떨어진다는 문제가 있다. 본 논문에서는 반복적 깊이 예측 구조의 개선을 통해 출력된 깊이맵들의 깊이 일관성을 높이는 방법을 제안하고자 한다.

Ⅰ. 서 론

최근 VR, AR, 메타버스 산업이 발전하면서, 다양한 방법으로 3D 재구성이 시도되고 있다. LiDAR, RGB-D 와 같은 특수한 센서를 사용할 경우 정확도는 높지만 비용이 높아 활용도가 떨어진다. 최근 Deep Learning 기술의 발전으로 이러한 깊이 추정을 위한 별도의 센서 없이 일반 카메라로부터 촬영된 RGB 이미지로부터 3 차원 장면을 얻어내는 3 차원 재구성 분야가 발전하고 있다.

Deep Learning 기반의 3 차원 재구성의 경우 깊이 이미지가 사용된다. 깊이 이미지란 카메라로부터의 물체까지의 거리를 표현한 이미지이다. 깊이 이미지를 역투영 (Back-Projection)하여 3 차원 영상으로 복원할 수 있다. 단일 RGB 이미지로부터 깊이 이미지를 추정하고 이를 입력 받아 3 차원 재구성을 하는 다양한 Deep Learning Model 이 있다. 이때 깊이 이미지의 품질이 3 차원 재구성의 완성도를 좌우하는 중요한 요소 중 하나이다. RGB 이미지 내에서 물체와 물체 간의 폐색은 3 차원 재구성 과정에서 다양한 어려움을 야기시킨다. 단일 RGB 이미지로부터 깊이를 예측할 경우 가리는 물체(Occluder) 뒤에 있는 가려진 물체(Occludee)의 일부 영역의 깊이는 추정하기 어렵다. 즉, 가려진 부분의 깊이 정보의 부족으로 인해 3 차원 재구성의 완성도 떨어지고 다양한 구멍(hole)이 관찰된다.

이러한 문제를 해결하기 위해 [1]에서는 폐색이 일어나는 레이어마다 개별적인 깊이를 추정하여 가려진 깊이 정보를 복원하는 반복적 깊이 예측 기법이 제안되었다. 반복적 깊이 예측 구조를 통해 폐색이 일어난 영역의 깊이가 잘 복원될 경우 3 차원 재구성에서도 해당 영역을 복원할 수 있게 된다. 그러나 반복적 깊이 예측의 경우 예측되는 깊이들이 균일하지 않을 수 있다는 문제가 있다. 본 논문에서는 반복적 깊이 예측 구조를 조정하여 깊이 일관성을 향상시키기 위한 방법을 제시한다.

Ⅱ. 본 론

본 논문은 3 차원 재구성을 위한 반복적 깊이 예측 시스템을 기반으로 한다 [1]. 반복적 깊이 예측 구조에서는 먼저 RGB 이미지로부터 폐색이 일어난 물체의 마스크를 추출한 후, 이를 활용해 최종적으로 깊이를 추정한다. 본 연구에서 사용된 데이터셋은 Front3D[3]를 기반으로 한 Amodal-3D-Front-Dataset 으로, 이는 3 개의 물체 간에 폐색이 발생한 실내 합성 데이터셋이다. 본 논문은 이러한 구조와 데이터셋을 활용하여 반복적 깊이 예측에서 발생하는 깊이 일관성 문제를 해결하는 새로운 기법을 제안한다.

반복적 깊이 예측 구조는 폐색이 발생할 때마다 깊이를 예측한다. 기존의 Amodal Depth Estimation (ADE) Model 은 Vision Transformer(ViT) 인코더와 Convolution 기반 Decoder 를 사용하며, Scale-shift invariant loss 를 통해 정규화된 깊이 이미지를 출력한다. 이 Model 의 입력은 단일 RGB 이미지와 폐색된 물체의 마스크(Amodal Mask)이다. 그림 1 (a)와 같이 현재 데이터셋에서는 3 개의 폐색 물체가 있으므로, 모델은 3 번의 깊이 출력 과정을 거친다. 그러나 이 구조는 반복적으로 출력된 깊이들이 각각 독립적으로 처리되어, 동일한 영역에 대한 일관성이 떨어지는 문제를 야기한다. 그림 1 (b)는 본 논문에서 제안하는 반복적 깊이 예측

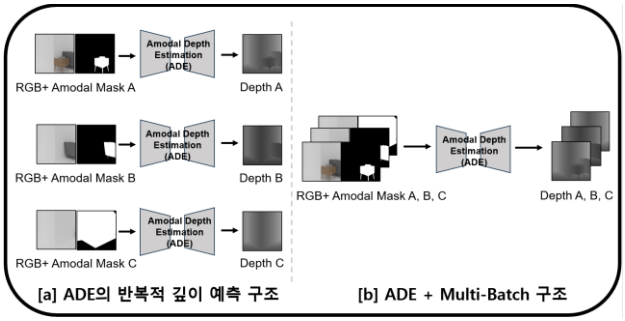

그림 1. Amodal Depth Estimation Model(ADE)의 변형 구조 (a) 기본 ADE Model 의 반복적 깊이 예측 (b) Multi-Batch 를 활용한 ADE Model 의 다중 깊이 예측

구조를 보여준다. 기존 구조를 개선하기 위해, 본 논문에서는 Amodal Depth Estimation 모델의 인코더에 폐색된 3 개의 Amodal Mask 를 Batch 단위에서 한 번에 입력하고, Decoder 에서 3 개의 깊이를 출력하는 ADE Multi-Batch 모델을 제안한다. 이를 통해 제안된 모델은 기존 ADE 모델에 비해 높은 품질의 깊이를 출력하며, 에이모달 깊이 이미지 간의 일관성도 향상시킬 수 있다.

Ⅲ. 실 험

본 논문에서는 기본 ADE 모델, 기본 ADE 모델에 Consistency Loss 를 추가한 모델, ADE 에 Multi-Head-Attention(MHA)를 추가한 모델, 그리고 ADE 에 Batch 단위로 3 장의 입력과 출력을 처리하는 Multi-Batch 모델을 비교하여 실험한다. 표 1 은 ADE+ Multi-Batch 모델이 출력한 Depth 의 정량적 평가를 나타낸다. 출력된 Depth 는 threshold accuracy δ^n (percent of pixels max($D_{inference}/D_{GT}$, $D_{GT}/D_{inference}$))와 root mean squared error(RMSE)로 측정한다. 기존 Model 과 비교했을 때, ADE+ Consistency Loss Model 은 δ^2, δ^3 에서 가장 높은 성능을 보이며, δ^1, RMSE 에서는 ADE+ Multi-Batch 가 가장 우수한 성능을 나타낸다. 이 결과는 제안된 모델이 기존 방법에 비해 깊이 추정 성능을 향상시켰음을 보여준다.

표 2 는 깊이 일관성을 정량적으로 평가한 결과를 보여준다. 깊이 일관성 향상은 3 차원 재구성에서는 중요한 과제이지만, 일반적으로 정적인 이미지에서는 거의 동일한 장면과 시점의 깊이 일관성은 크게 고려되지 않는다. 그러나 본 논문에서 고려하는 구조와 데이터셋은 같은 시점에서 객체가 이동하는 듯한 형태를 가지며, 이는 Video 와 유사하다. 따라서 깊이 일관성을 측정하기 위해 전통적인 Video Depth Estimation 분야의 Metric 들을 적용한다. 이러한 Metric 에는 출력된 Depth Map 을 투영하여 만든 Point Cloud 의 일치율을 기반으로 하는 Fitness, Point Cloud RMSE 등이

표 1. 출력된 Depth 의 정량적 평가

	δ^1	δ^2	δ^3	RMSE
ADE	0.98366	0.99621	0.99907	0.13641
ADE + Consistency Loss	0.98367	**0.99624**	**0.99913**	0.13634
ADE + Multi-Head Attention	0.93419	0.98504	0.99632	0.27293
ADE + Multi-Batch	**0.98373**	0.99613	0.99907	**0.13601**

표 2. 출력된 Depth 간의 Consistency 측정

	Pixel-wise Depth Difference	Depth Histogram Comparison	Fitness	Point Cloud RMSE
ADE	0.22461	0.73115	0.48614	0.01013
ADE + Consistency Loss	0.22354	0.72514	0.48124	0.01005
ADE + Multi-Head Attention	0.31577	0.64568	0.19311	0.01137
ADE + Multi-Batch	**0.21884**	**0.73616**	**0.51720**	**0.00992**

포함된다. 이 외에도 Depth Map 의 픽셀 값을 기반으로 일치율을 계산하는 Pixel-wise Depth Difference, Depth Histogram Comparison 등을 통해 Depth Consistency 를 측정하였다. 측정된 모든 Depth Consistency Metric 에서 제안된 Model 이 가장 높은 성능을 보였으며, 이를 통해 제안된 방법으로 출력된 Depth 간의 일관성이 향상되었음을 확인할 수 있다. 이러한 결과는 ADE 모델의 Encoder 와 Decoder 를 변경함으로써 깊이의 품질뿐만 아니라 깊이 간의 일관성도 향상시킬 수 있음을 시사한다.

Ⅳ. 결 론

본 논문에서는 폐색 환경의 반복적 깊이 예측 구조에서 Encoder, Decoder 를 개선하여 깊이 추정에서의 일관성 및 품질을 향상하는 방법을 제시한다. 이러한 접근을 통해 단일 이미지에서의 3 차원 재구성의 품질을 높이는데 기여할 수 있기를 기대한다.

ACKNOWLEDGMENT

This research was supported by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2021-0-02052) supervised by the IITP (Institute for Information & Communications Technology Planning & Evaluation)

참 고 문 헌

[1] Seung-Uk Jo, Du Yeol Lee, Chae Eun Rhee, “Occlusion-aware Amodal Depth Estimation for Enhancing 3D Reconstruction from a Single Image”, in Revision of IEEE Access, 2024

[2] Hyunmin Lee and Jaesik Park. Instance-wise occlusion and depth orders in natural scenes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 21210-21221, 2022.

[3] H. Fu, B. Cai, L. Gao, L.-X. Zhang, J.Wang, C. Li, Q. Zeng, C. Sun, R. Jia, B. Zhao et al., “3d-front: 3d furnished rooms with layouts and semantics,” in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021, pp. 10 933-10 942.