최근접 이웃 커널 추정을 통한 희소 깊이 영상 완성 네트워크

정태현, 오병태
한국항공대학교 항공전자정보공학부
sdfds5566@kau.kr, byungoh@kau.ac.kr

Sparse Depth Image Completion Network with nearest neighbor kernel estimation

TaeHyun Jeong, Byung Tae Oh
Korea Aerospace University

요 약

본 논문에서는 희소깊이영상과 컬러영상을 이용해 조밀한 깊이영상을 추정하는 깊이 완성(depth completion)을 수행하기위해 최근접 이웃 커널을 추정하는 방식의 네트워크를 제안한다. 회귀방식의 딥러닝 네트워크는 일반적으로 값을 직접 예측하는 것보다 기본 값에 더해질 잔차를 추정하는 방식이 더욱 효율적이다. 본 논문에서는 최근접 이웃 커널을 입력영상에 적용하여 추정하고자 하는 픽셀의 인근 픽셀에서 값을 가져와 기본 값으로 사용하고, 해당 값의 잔차를 회귀방식으로 추정하는 네트워크를 설계했다. 이러한 방식으로 여러 SOTA 알고리즘 대비 좋은 성능을 나타냈고, 특히 이와 유사한 방식인 Plane-residual net 보다 높은 성능을 보여준다.

1. 서론

최근 영상 센싱 기술의 발전으로 영상을 표현하는 기존의 RGB 영상 데이터와 다른 형태의 데이터를 이용해 시각화 혹은 처리하는 방식이 활발하게 연구되고 있다. 그 중에서도 특히 ToF, LiDAR 등의 거리측정센서로 얻을 수 있는 깊이 영상(depth map)과 같은 3 차원 공간정보를 가진 데이터에 대한 처리기술은 자율주행, 3 차원 영상, VR 등 각종 산업군에서 반드시 필요한 기술이다.

깊이 영상은 하드웨어 특성상 컬러영상만큼 높은 해상도의 데이터를 획득하기 어렵고, 어떤 픽셀에서 유효한 데이터가 관측될지 특정할 수 없다는 특징이 있다. 이러한 특성을 가진 희소 깊이 영상(sparse depth map)의 관측되지 않은 비유효 픽셀들을 추정하여 컬러영상과 같은 해상도의 조밀한 깊이 영상(dense depth map)을 획득하는 과업을 깊이 완성(depth completion)이라고 한다. 일반적으로 깊이완성을 수행하기 위한 딥러닝 네트워크는 희소깊이영상 뿐만 아니라 텍스쳐 정보를 학습하기 위해서 컬러영상 또한 입력으로 사용하고, 조밀한 깊이 영상을 출력하는 방식으로 설계된다. 네트워크 구조는 주로 깊이 영상에 대한 충분한 특징을 학습하기 위해 인코더-디코더 구조를 사용한다.

깊이 완성을 수행하기위한 다양한 접근 방식이 있다. 먼저 희소영상의 특성을 잘 학습하기 위해 희소 영상에 대응하는 이진 마스크를 은닉층에서 max pooling 을 통해 전달하는 sparsity-invariant network [1]와 이 방식을 응용하여 이진 마스크를 신뢰도 맵으로 간주하여 은닉층에서 주변으로 전파되는 방식의 컨볼루션 연산을 고안한 NCONV [2]가 있다. 또한 컬러 영상과 깊이 영상의 특징맵을 융합시키는 방식에 초점을 맞추어 좋은 성능을 낸 네트워크들도 있다. Guide net [3]은 컬러 영상으로 만든 특징맵을 이용해 동적 필터를 생성하고, 희소 깊이 영상 네트워크에 적용하는 방식이다. FCFR-Net [4]은 희소 깊이영상

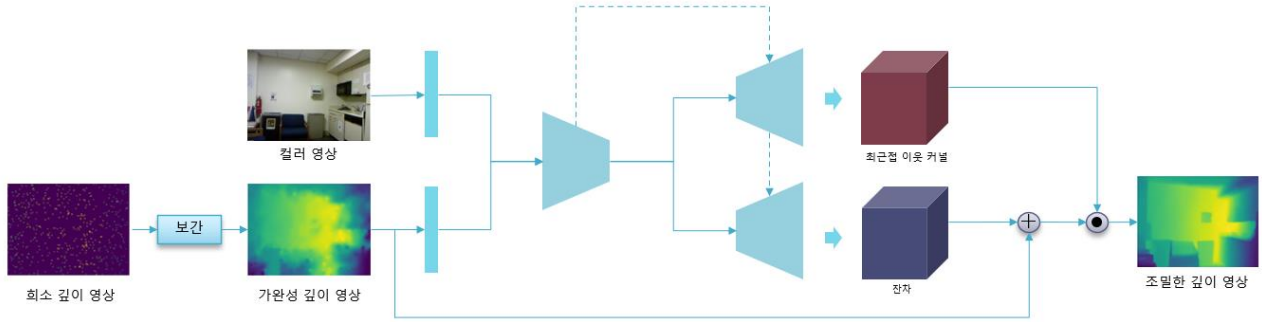

그림 1 제안 모델의 구조도

대신 가완성된 깊이 영상(pseudo depth map)을 입력 받고, 컬러영상의 특징맵과 깊이 영상의 특징맵을 채널 셔플과 에너지기반의 융합을 통해 두 영상간의 관계를 풍부하게 학습 할 수 있다. 픽셀값을 주변으로 전파시키는 방식의 네트워크들도 있다. SPN [5]은 주변 픽셀과의 관계를 정의하는 affinity 를 추정하여 영상에 적용한다. CSPN [6]은 SPN 에서 대각선 방향을 추가하고, affinity 를 커널 형태로 만들어 영상에 적용하는 방식으로 설계되었다. 데이터의 표현 방식을 바꿔 훈련시킨 네트워크도 있다. Plane-residual net [7] 은 깊이 영상의 값을 기준으로 n 개의 plane 으로 분할한 plane 영상과 해당 plane 에 대한 잔차 영상으로 표현하여 네트워크를 학습시킨다. 각 픽셀마다 어떤 plane 에 속하는지는 분류 손실 함수를 적용하여 찾고, 잔차는 회귀 손실 함수를 적용하여 찾는다. 해당 네트워크는 이러한 방식으로 각 픽셀을 plane 으로 대략적으로 찾은 후 디테일은 잔차로 찾는 방식이다. 회귀해야할 값의 범위가 작아지기 때문에 네트워크의 부담이 줄고 효율적인 방식이다.

본 논문은 [7]의 방식에서 영감을 얻어 plane 대신 인근의 값을 가져오고, 잔차를 더해주는 방식의 네트워크를 고안했다. 비유효 픽셀에서 유효한 픽셀값을 가져옴으로써 plane-residual net 보다 잔차의 추정 범위가 작아지기 때문에 더욱 좋은 성능을 낼 수 있다.

2. 제안하는 기법

제안하는 네트워크는 컬러 영상과 희소 깊이영상으로부터 보간된 가완성 깊이영상을 입력받아 조밀한 깊이영상을 직접 추정하는 것이 아닌 최근접 이웃 커널과 잔차를 추정한다. 백본 네트워크는 resnet-18 을 사용하였고, 그림 1 과 같이 컬러영상과 깊이영상 정보를 하나의 인코더로 추출하고, 해당 정보를 바탕으로 두개의 디코더를 통해 각각 커널과 잔차를 추정한다.

2.1 최근접 이웃 커널

최근접 이웃 커널은 dynamic local filter [8]와 같은 형태로 $k^2 * H * W$의 형태를 가진다. H, W는 입력 이미지의 가로, 세로 크기이고, k^2은 각각의 픽셀마다 적용될 커널의 사이즈이다. 해당 커널은 채널 방향으로 softmax 를 취하여 $k \times k$ 영역의 픽셀에 대한 가중치를 나타낸다. 이후 입력된 가완성 깊이 연산에 가중합계 연산으로 적용되어 이웃한 픽셀 중 정답에 가까운 픽셀에 좀 더 가중치를 두어 연산이 진행된다.

2.2 잔차 추정

잔차도 최근접 이웃커널과 동일한 형태를 가진다. 최근접 이웃 커널을 적용하기 전에 먼저 커널 영역에 잔차를 더해준다. 하나의 픽셀이 아닌 커널 전체 영역에 잔차를 더해줌으로써 최근접 이웃 커널의 예측 안정성을 향상시킬 수 있다. 또한 각 픽셀에 적용되는 커널영역마다 적응적으로 잔차를 더해주기 때문에 최근접 이웃 커널의 가중치 선택에 도움을 줄 수 있다.

2.3 희소 깊이영상 보간

희소 깊이영상에 커널을 그대로 적용할 경우, 커널영역에 유효픽셀이 없는 경우가 존재한다. 따라서 입력 희소영상을 모델기반의 보간법을 사용해 대략적으로 추정한다. 보간의 최종단계에서 입력희소영상이 원래 가지고 있었던 값은 유지한다. 이렇게 얻은 가완성된 깊이영상은 그림 1 과 같이 품질이 매우 떨어지지만, 잔차에 의해 보정될 수 있다.

3. 실험

본 논문의 깊이완성 네트워크를 평가하기위해 NYUv2 데이터셋을 사용하였다. NYUv2 는 마이크로소프트의 Kinect 를 통해 얻은 여러 실내영상의 RGB-D 데이터셋이다. NYUv2 는 48k 개의 훈련데이터 셋과 654 개의 테스트 데이터셋으로 구성되고, RGB 영상은 640x480 의 사이즈를 가지나 깊이 영상과 해상도를 맞추기 위해 다운샘플링후 center crop 을 통해 304x228 사이즈로 변환한다. 데이터셋의 깊이영상의 경우 정답 영상 전체 픽셀에서 500 개에 해당하는 비율로 랜덤

표 1 성능 비교

모델	RMSE(m)	REL(m)
NConv	0.129	0.018
CSPN	0.117	0.016
FCFR-net	0.106	0.015
Plane-residual net	0.104	0.014
Guide net	0.101	0.015
제안 모델	**0.100**	**0.013**

샘플링하여 희소 깊이영상을 만든다.

일반적으로 깊이영상을 평가할 때 RMSE 와 REL(mean absolute relative error)를 사용한다. RMSE 는 $\sqrt{\frac{1}{|D|}\sum_{d\in D}\|d^* - d\|^2}$, REL 은 $\frac{1}{|D|}\sum_{d\in D}|d^* - d|\,/d^*$ 로 표현된다. d 는 네트워크가 예측한 값, d^* 는 정답값, $|D|$ 는 전체 픽셀의 개수를 의미한다. 손실함수는 L1 loss 를 사용하였다.

표 1 은 여러 깊이 완성 모델들과의 성능을 비교하기 위해 도시한 결과이다. 여러 SOTA 모델들과 비교하여 우수한 성능을 보였고, 본 논문과 유사하게 잔차의 범위를 줄여 픽셀값을 추정하는 Plane-residual 방식 대비 약 4%의 성능을 향상 시켰다.

4. 결론

본 논문은 깊이 완성을 위해 최근접 이웃 커널을 추정하여 입력 깊이 영상에 적용하는 방식을 제안하였다. 회귀 추정할 잔차의 범위를 줄이기위해 인근 값에 잔차를 더하고 해당 픽셀값에 근접한 값을 가져오는 커널을 추정하는 방식으로 여러 SOTA 모델과 더불어 연구의 동기가 된 Plane-residual net 보다 우수한 성능을 보였다.

감사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단 기초연구사업(NRF-2022R1A2C1005769)과 경기도 지역협력연구센터 사업 (GRRC) (2017-B02, 3 차원 공간 데이터 처리 및 응용기술 연구)의 지원을 받아 수행되었음.

참조문헌

[1] Uhrig, Jonas, et al. "Sparsity invariant cnns." 2017 international conference on 3D Vision (3DV). IEEE, 2017.
[2] Eldesokey, Abdelrahman, Michael Felsberg, and Fahad Shahbaz Khan. "Confidence propagation through cnns for guided sparse depth regression." IEEE transactions on pattern analysis and machine intelligence 42.10 (2019): 2423-2436.
[3] Tang, Jie, et al. "Learning guided convolutional network for depth completion." IEEE Transactions on Image Processing 30 (2020): 1116-1129.
[4] Liu, Lina, et al. "FCFR-Net: Feature fusion based coarse-to-fine residual learning for depth completion." arXiv preprint arXiv: 2012.08270 (2020).
[5] Liu, Sifei, et al. "Learning affinity via spatial propagation networks." Advances in Neural Information Processing Systems 30 (2017).
[6] Cheng, Xinjing, Peng Wang, and Ruigang Yang. "Learning depth with convolutional spatial propagation network." IEEE transactions on pattern analysis and machine intelligence 42.10 (2019): 2361-2379.
[7] Lee, Byeong-Uk, Kyunghyun Lee, and In So Kweon. "Depth Completion using Plane-Residual Representation." Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
[8] Jia, Xu, et al. "Dynamic filter networks." Advances in neural information processing systems 29 (2016).