

반복적인 격자 워핑 기법을 이용한 깊이 영상 초해상도 기술

양윤모, 김동신, 오병태

한국항공대학교

yym064@kau.kr, dkim20@kau.kr, byungoh@kau.ac.kr

Iterative Deep Convolutional Grid Warping Network for Joint Depth Upsampling

Yoonmo Yang, Dongsin Kim, Byung Tae Oh
Korea Aerospace University

요 약

본 논문에서는 딥러닝 기반의 깊이 영상 초해상도 기술에 대해서 제안한다. 기존 깊이 영상의 초해상도 기술은 고해상도의 컬러 영상과 저해상도 깊이 영상을 이용하여 화소 값을 개선시켜 고해상도의 깊이 영상을 예측하였다. 하지만 이러한 방법들은 단순히 화소 값을 증가 또는 혹은 감소시키는 방법으로 언더슈팅 또는 오버슈팅과 문제를 발생시켜 성능 향상을 제한한다. 제안하는 기법에서는 이러한 한계를 극복하기 위해 화소의 위치를 이동하여 영상을 복원하는 격자 워핑 방식을 반복적으로 적용하여 고해상도 깊이 영상을 예측하였다. 실험 결과, 제안한 방식이 기존 방법들에 비해 정량적, 시각적 품질을 개선시켰음을 확인하였다.

This paper proposes a novel deep learning-based method to upsample a depth map. Most conventional methods estimate high-resolution depth map by modifying pixel value of given depth map using high-resolution color image and low-resolution depth map. However, these methods cause under- or over- shooting problems that restrict performance improvement. To overcome these problems, the proposed method iteratively performs grid warping scheme which shifts pixel values to restore blurred image for estimating high-resolution depth map. Experimental results show that the proposed method improves both quantitative and visual quality compared to the existing method.

1. 서론

최근 자율주행 및 로봇 공학 기술이 주목받음에 따라 3D 기술에 대한 관심이 증가하고 있다. 특히 3D 기술 중 깊이 정보를 획득하고 처리하는 기술에 대한 요구가 점점 더 중요해지고 있다. 깊이 정보를 획득하는 방법에는 여러가지가 있다. 이 중 대표적으로 많이 사용하는 방식이 수동적 방식과 능동적 방식을 예로 들 수 있다.

수동적 방식은 깊이 정보를 간접적으로 획득하는 방식으로 스테레오 정합 방식이 수동적 방식에 대표적인 예라고 할 수 있다. 스테레오 정합방식은 여러 장의 사진에서 서로 매칭되는 위치를 찾아 깊이 정보를 추출하는 방식으로 매칭되는 위치를

찾는데 많은 연산을 요구한다. 또한 반복되는 값을 갖는 지역 또는 폐쇄 영역에서 큰 오차를 발생하는 특징이 있다.

능동적인 방식은 레이저 거리계, Time of Flight (TOF), 구조광 카메라 등과 같이 특별한 장치를 이용하여 깊이 정보를 직접적으로 획득하는 방식이다. 이러한 방식은 수동적 방식에 비해 비교적 빠른 시간안에 영상을 획득할 수 있다는 장점이 있지만, 현재 하드웨어 기술에 한계로 인해 해상도의 품질이 떨어지는 깊이 영상을 얻는다는 단점이 있다. 이러한 단점을 극복하기 위해 저해상도 깊이 영상의 해상도를 높이기 위한 연구들이 제안되고 있다. 과거에는 이러한 문제를 해결하기 위해 주로 최적화 기법 또는 결합 (joint) 필터 방식을 이용해왔다.

최근 딥러닝 (Deep Learning) 기술의 도래와 더불어 객체분류, 객체 검출, 이미지 분할 등의 분야에서 주목할 만한 성과를 보이고 있다. 이에 따라 깊이 영상의 해상도를 높이기 위해 딥러닝을 이용한 방법들이 제안되고 있다. 하지만 제안되고 있는 방법들은 단순히 기존 초해상도 (Super-resolution) 분야에서 사용되고 있는 기술을 차용하여 깊이 영상에 해상도를 높이기 위해 기술을 확장시킨 방법으로 성능 향상에 한계점이 존재한다. 본 논문에서는 기본 방식들의 한계를 극복하기 위해 딥러닝 기반의 격자 워핑 방식을 제안하였다 [1].

2. 제안 알고리즘

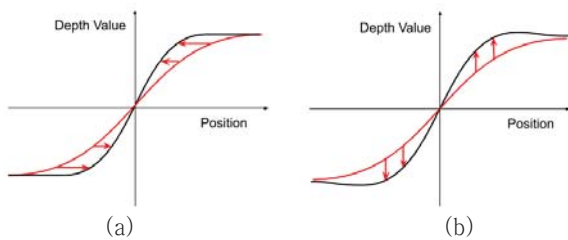


그림 1. 블러된 화소 값(빨강)을 블러되지 않은 화소 값(검정)로 복원하는 방법: (a) 픽셀 위치 이동 (b) 픽셀 값 보상

기존 방법들을 살펴보면 그림 1에 (b)와 같이 공간적 정보를 참조하여 픽셀 값을 증가 혹은 감소시켜 고해상도의 깊이 영상을 예측하는 방식이다. 하지만 이런 방법들의 문제점은 그림에서 보는 것과 같이 언더슈팅 또는 오버슈팅을 유발하여 성능 향상에 한계를 보인다. 제안하는 방식은 기존 방식과 다르게 저해상도의 영상을 bicubic 보간법을 이용하여 업샘플링하고, 업샘플링 과정에서 발생하는 블러링 왜곡을 최소화하여 고해상도의 깊이 영상을 예측하였다.

이때 블러링 왜곡을 최소화하기 위해 격자 워핑 방식을 사용하였다. 전통적인 격자 워핑 방식은 가우시안 블러와 같이 균일한 블러링 왜곡(Uniform blurring error)를 최소화하도록 설계되었다. 블러링 왜곡은 블러링 과정에서 경계 값 주변의 픽셀 값들을 경계 바깥쪽으로 이동시킴으로써 왜곡이 발생한다고 가정한다. 따라서 왜곡을 복원하기 위해서는 그림 1의 (a)와 같이 경계 바깥쪽으로 이동된 픽셀들을 경계 중심 쪽으로 이동시킴으로써 경계를 복원한다. 이때 기존 격자 워핑 방식은 블러링 왜곡이 포함된 영상의 정보만을 이용하여 픽셀 값을 이동시키는 변위 벡터를 유도한다. 변위 벡터를 유도하는

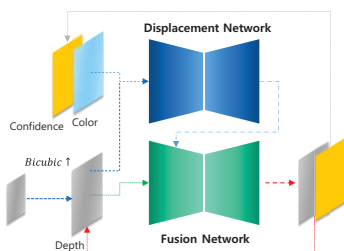


그림 2. 제안 시스템의 순서도.

과정에서 경계의 중심 위치에 따라 변위 벡터의 크기와 방향이 결정이 되는데, 제안하는 시스템에서는 변위 벡터를 추출하기 위해 블러링 왜곡이 포함된 깊이 영상과 함께 고해상도 컬러 영상을 이용하여 변위 벡터를 도출하도록 시스템을 구성하였다.

또한 제안하는 시스템에서는 고해상도 깊이 영상과 영상의 신뢰도를 함께 예측하고 이를 기반으로 반복적으로 고해상도 깊이 영상을 예측하도록 시스템을 구성하였다. 신뢰도를 기반으로 깊이 영상을 반복적으로 예측하는 이유는 깊이 영상의 경계영역에서 경계의 중심 위치를 보다 정확하게 예측하여 더 정확한 변위 벡터를 유도하여 성능을 향상시키기 위해서이다.

시스템은 구성의 그림 2 와 같다. 시스템은 크게 2 가지 네트워크로 구성된다. Displacement network 는 각 화소의 위치에서 변위벡터를 예측하는 네트워크로 FlowNetS 과 같은 구조를 가지고 있다 [2]. Fusion Network 는 displacement network 에서 변위벡터를 받아 워핑을 수행한 피쳐맵을 정제하는 네트워크로, GridNet 과 같은 구조를 가지고 있다 [3].

3. 실험

제안하는 방법을 실험하여 검증하는데 있어서 각각의 배율을 따로 학습하였다. 시스템을 학습하기 위해 Adam optimizer 를 사용하였다 [4]. 학습률은 $1e^{-4}$ 으로 초기화 하였으며 매 5 번째 epoch 마다 학습률을 절반으로 감소시켜 학습을 진행하였다. 실험환경은 PyTorch 를 이용하여 구현하였다. 반복횟수는 출력 영상의 품질과 네트워크의 복잡도를 고려하여 3 번으로 설정하였다.

Dataset	Lu		NYU v2		Sintel		Middlebury	
	4x	8x	4x	8x	4x	8x	4x	8x
PAC[9]	1.39	3.42	1.52	2.75	3.72	5.42	1.61	3.54
DKN[10]	1.14	2.42	1.08	1.91	2.92	4.55	1.32	2.35
GWN[11]	1.04	2.24	0.91	1.60	2.66	3.81	1.27	2.07
Ours	0.91	1.67	0.80	1.37	2.46	3.12	1.23	1.84

표 1. RMSE 성능 비교

시스템을 학습하기 위해 Sintel, NYU 데이터셋 중 일부를 학습 데이터 셋으로 사용하였다 [5, 6]. 검증용 위해 Sintel

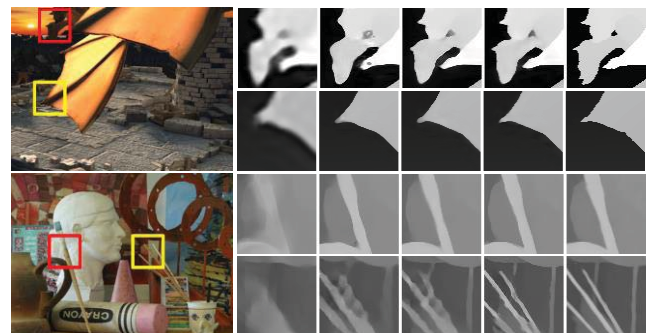


그림 3. 8배 업샘플링한 환경에서 Temple2와 Art 깊이 영상의 시각적인 비교: 왼쪽부터 PAC[9], DKN[10], GWN[11], 제안기술, 원본

NYU, Lu, Middlebury 데이터 셋을 이용하였다 [5-8].

제안 기술을 평가하기 위해 기존 딥러닝 기반 깊이 영상 업샘플링 방법들과 root mean square error(RMSE)를 평가 지표로 사용하여 성능을 비교하였다 [9-11]. 표1에서 보이는 것과 같이 제안하는 방법이 모든 데이터 셋에서 가장 좋은 성능을 보이는 것을 확인할 수 있다.

그림 3 에서 보는 것과 같이 제안하는 방법은 경계정보를 가장 잘 복원하는 것을 확인할 수 있다. 그림 3 에 두번째 줄에서 볼 수 있듯 제안하는 방법은 다른 방법과 비교해 원본에 가까운 형태를 잘 복원하는 것을 확인할 수 있다. 특히 그림 3 에 네번째 줄과 같이 bicubic 보간법을 수행하는 과정에서 생긴 블러링 왜곡으로 인해 물체가 겹쳐지는 복원이 어려운 경우에도 물체를 비교적 잘 분리하여 좋은 성능을 보이는 것을 확인할 수 있다.

4. 결론

본 논문에서는 반복적 격자 워핑 기법을 통한 깊이 영상의 초해상도 기법에 대해서 제안하였다. 제안하는 방법은 고해상도 컬러 영상과 저해상도 깊이 영상의 정보들을 통하여 변위 벡터를 유도하는 시스템을 제안하였고 신뢰도를 기반으로 격자 워핑을 반복적으로 적용하는 시스템을 설계하여 성능을 크게 향상시켰다. 하지만 제안하는 방법은 반복적인 연산으로 인해 많은 연산량을 요구한다는 한계점이 존재한다. 추후 연구에서는 이러한 단점을 극복하는 방향으로 연구를 진행할 예정이다.

감사의 글

본 연구는 2020 년도 정부(교육부)의 재원으로 한국연구재단 기초연구사업(NRF- 2019R1F1A1063229)과 경기도 지역 협력 연구센터 사업 (GRRC) (2020-B02, 3 차원 공간 데이터 처리 및 응용기술 연구)의 지원을 받아 수행되었음.

참조문헌

- [1] A. Nasonova and A. Krylov, "Deblurred images post-processing by Poisson warping," IEEE Signal Process. Lett., vol. 22, no. 4, pp. 417-420, Apr. 2015.
- [2] A. Dosovitskiy, P. Fischer, E. Ilg, "Flownet: Learning optical flow with convolutional networks," IEEE Computer Vision and Pattern Recognition, pp. 2758-2766, 2015.
- [3] D. Fourure, R. Emonet, E. Fromont, D. Muselet, A. Tremeau, C. Wolf, "Residual conv-deconv grid network for semantic segmentation." arXiv preprint arXiv:1707.07958, 2017.
- [4] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980, 2014.
- [5] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black, "A naturalistic open source movie for optical flow evaluation," in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, pp. 611-625, 2012.
- [6] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," in Proc. Eur. Conf. Comput. Vis. Cham, Switzerland: Springer, pp. 746-760, 2012.
- [7] H. Hirschmuller and D. Scharstein, "Evaluation of cost functions for stereo matching," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 1-8, Jun. 2007.
- [8] S. Lu, X. Ren, and F. Liu, "Depth enhancement via low-rank matrix completion," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., pp. 3390-3397, Jun. 2014.
- [9] H. Su, V. Jampani, D. Sun, O. Gallo, E. Learned-Miller, and J. Kautz, "Pixel-adaptive convolutional neural networks," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., pp. 11166-11175, Jun. 2019.
- [10] K. Beomjun, J. Ponce, H. Bumsu, "Deformable kernel networks for guided depth map upsampling," arXiv preprint, arXiv:1910.08373, 2019.
- [11] Y. Yoonmo, K. Dongsin, O. Byung Tae, "Deep Convolutional Grid Warping Network for Joint Depth Map Upsampling." IEEE Access, vol. 8, pp. 147580-147590, Aug 2020