

에지 맵을 이용한 CNN 기반 단일 영상의 깊이 추정

고영권, 문현철, 김현호, 김재곤

한국항공대학교

{dram109, hcmoon, hhkim}@kau.kr, jgkim@kau.ac.kr

Single-Image Depth Estimation Based on CNN Using Edge Map

Yeong-Kwon Ko, Hyeon-Cheol Moon, Hyun-Ho Kim, and Jae-Gon Kim
Korea Aerospace University

요 약

CNN(CNN: Convolutional Neural Network)은 컴퓨터 비전의 많은 분야에서 뛰어난 성능을 보이고 있으며, 단일 영상으로부터 깊이(depth) 추정에서도 기존 기법보다 향상된 성능을 보이고 있다. 그러나, 단일 영상으로부터 신경망이 얻을 수 있는 정보는 제한적이기 때문에 스테레오 카메라로부터 얻은 좌/우 영상으로부터의 깊이 추정보다 성능 향상에 한계가 있다. 따라서 본 논문에서는 에지 맵(edge map)을 이용한 CNN 기반의 단일 영상에서의 깊이 추정의 개선 기법을 제안한다. 제안 방법은 먼저 단일 영상에 대한 전처리를 통해서 에지 맵과 양방향 필터링된(bilateral filtered) 영상을 생성하고, 이를 CNN 입력으로 하여 기존 단일 영상 깊이 추정 기법 대비 개선된 성능을 보임을 확인하였다.

1. 서론

심층 신경망은 컴퓨터 비전의 많은 분야에서 뛰어난 성능을 보이고 있으며, 단일 영상으로부터 깊이 추정에서도 기존의 기법보다 향상된 성능을 보이고 있다. 예를 들어, Eigen[1] 등의 방식에서는 합성곱 신경망(Convolutional Neural Network: CNN)에 단일 영상을 입력 받아 깊이 지도(depth map)를 생성하고 해당 결과를 바탕으로 지역 영역을 개선하는 두 단계(Two-stage)의 심층 신경망 구조를 제안했다. 하지만 신경망 구조상 풀링(Pooling) 연산을 사용하면서 출력 영상의 해상도가 감소함과 동시에 화소 정보가 손실되고, 따라서 객체의 경계선 등 지역적 영역 내 상세 정보가 온전히 생성되지 못하는 문제가 발생한다. 또한 단일 영상으로부터 CNN 이 얻을 수 있는 정보는 제한적이기 때문에 스테레오 카메라로부터 얻은 좌/우 영상으로부터의 깊이 추정보다 성능 향상에 한계가 있다.

본 논문에서는 기존의 단일 영상에서의 깊이 추정 기법에서 입력 영상으로부터 얻을 수 있는 정보가 제한적인 것에 주목했다. 기존의 DE-CNN[2]의 방식에서 영상을 전 처리하여 기존의 깊이 영상과 같이 신경망의 입력으로 하여 깊이 영상을 개선시키고자 하는 시도가 있었다. 이를 깊이 추정에 적용하여 영상의 에지 맵(edge map)을 이용한 CNN 기반의 단일 영상에서의 깊이 추정의 개선 기법을 제안한다. 즉, 기존의 입력 영상만을 입력을 하는 것과 달리 에지 맵과 양방향(bilateral) 필터링된 영상을 CNN 의 입력으로 하여 깊이 추정을 하는 방식을 제안한다.

2. 제안 깊이 추정 기법

기존의 CNN 을 활용한 깊이 추정 방식의 성능 향상을 위해 입력 영상에 대한 전처리를 수행했다. 기존의 깊이 영상의 경우 에지 특성이 제대로 표현되지 않는 경우가 많아 이를 보완하기 위해 입력 영상으로부터 에지 맵(edge map)을 추출하여 CNN 의 입력으로 한다. 추가적으로, 입력 영상의 상세한 텍스처(texture) 특징이 불필요하게 깊이 영상에 강하게 나타나는 것을 막기 위해서 입력 영상의 에지 성분을 보존하면서 평활화(smoothing)를 할 수 있는 양방향(bilateral) 필터로 필터링된 영상을 CNN 의 입력으로 사용한다. 그림 1은 입력 예를 보인 것이다.

제안 기법에서 CNN 구조는 U-Net[3] 기반의 그림 2 와 같은 구조를 사용한다. U-Net 구조는 Encoder-Decoder 형태로 구성되고 Encoder 는 문맥(context)를 획득하는 역할을 하고 대칭적인 구조를 이루는 Decoder 에서는 정교한 지역화(localization)를 가능하게 하는 구조이다. 이러한 특징으로 많은 비주얼 인식작업에서 매우 좋은 성능을 보여주고 있어 깊이 추정에도 적절한 구조로 판단된다.

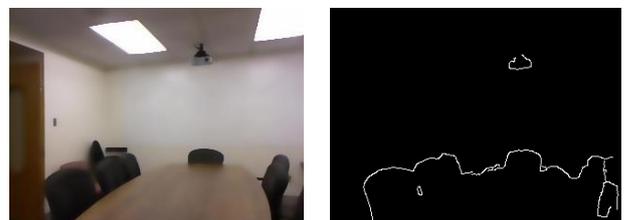


그림 1. 양방향 필터링 이미지와 에지 맵 예

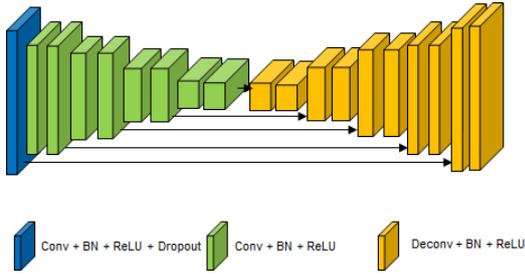


그림 2.U-Net 구조 기반의 깊이 추정 네트워크 구조

3. 실험결과

실험 데이터셋으로 실내 환경에 대한 깊이 영상을 제공하는 NYU Depth v2[4]의 Raw Version 을 사용했다. NYU Depth v2 은 키넥트 카메라를 이용하여 촬영한 464 개의 장면의 깊이 영상을 제공한다. 그중 249 개 장면을 학습에 사용하고 215 개 장면을 테스트에 사용하였다. 해상도는 640x480 이며 본 실험에서는 320x240 해상도를 사용했다. 학습과 테스트 데이터셋으로 각각 100K 개와 694 개의 영상을 사용하였다. 학습 손실함수(Loss Function) L1, L2, Huber Loss 에 대해서 실험을 진행했고 최종적으로 가장 좋은 성능을 보인 L1(Mean Absolute Error)를 선택하였다.

기존의 깊이 추정 방식들의 일반적인 성능평가 방식[5], [6]을 사용해서 NYU 데이터셋 깊이 영상의 Ground Truth 와 CNN 의 깊이 추정 영상을 비교하였다. 제안된 기법의 성능평가를 위해 U-Net[3] 구조에 단일 입력 영상을 입력으로 하는 결과와 해당 구조에 제안한 기법을 적용한 결과를 비교하였다.

표 1 은 기존의 단일 영상 입력 결과와 제안된 기법의 성능 비교를 나타낸다. 추정된 깊이의 정확도와 오류율의 전반적인 부분에서 성능향상을 확인할 수 있다. 표 2 는 기존의 기법과 제안 기법의 주관적 화질을 비교한 것이다. Best Case 는 기존 기법 대비 제안된 기법의 성능이 제일 좋은 경우이고, Worst Case 는 그 반대의 경우이다. 두 경우 모두 제안한 방식의 추정 깊이 영상(d)에서 객체나 배경의 에지 부분들이 보다 선명하게 표현되는 것을 확인할 수 있다

표 1. 제안된 기법의 실험결과

	U-net	U-net + Proposed	
$\delta < 1.25^1$ [6]	0.575	0.573	Higher is better
$\delta < 1.25^2$ [6]	0.812	0.819	
$\delta < 1.25^3$ [6]	0.923	0.927	
Absolute Relative Error [5]	0.301	0.289	Lower is better
Squared Relative Error [5]	0.283	0.262	
RMSE (linear) [5]	0.616	0.611	
RMSE (log) [5]	0.201	0.200	

4. 결론

본 논문에서는 기존의 CNN 기반의 단일 영상에서의 깊이 추정 기법을

개선하기 위하여 에지 맵과 양방향 필터링된 영상을 입력으로 사용하는 깊이 추정 기법을 제안하였다. 실험결과 제안 기법은 주관적 성능 지표 및 객관적 화질 비교에서 기존 기법 보다 향상된 성능을 보임을 확인하였다. 특히, 주관적 화질에서 객체나 배경의 윤곽선이 보다 선명하게 표현되는 것을 확인하였다.

표 2 주관적 화질 성능 비교(a)입력 영상, (b)Ground Truth, (c) 기존 CNN 기반 깊이 영상, (d)제안한 CNN 기반 깊이 영상

	Best Case	Worst Case
(a)		
(b)		
(c)		
(d)		

감사의 글

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원을 받아 수행된 연구임(No. 2020R1F1A10 68106)

참고 문헌

- [1] D. Eigen, C. Puhrsch, and R. Fergus, "U-net: Convolutional networks for biomedical image segmentation," In Proc. MICCAI 2015, 2015.
- [2] X. Zhang and R. Wu, "Fast depth image denoising and enhancement using a deep convolutional network," In Proc. ICASSP 2016, 2016.
- [3] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," Proc. MICCAI 2015, 2015.
- [4] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor segmentation and support inference from RGBD images," In Proc. ECCV, 2012.
- [5] C. Cadena, Y. Latif and I. D. Reid, "Measuring the performance of single image depth estimation methods," IROS 2016, 2016.
- [6] M. P. Lubor Ladicky, Jianbo Shi. "Pulling things out of perspective," In Proc. CVPR, 2014.