합성곱 신경망을 이용한 깊이맵 생성

김홍진 김만배

강원대학교 컴퓨터정보통신공학과

*solomoon1007@kangwon.ac.kr, manbae@kangwon.ac.kr

Depth map generation using convolutional neural network

Hong-Jin Kim and Manbae Kim

Computer and Communications Engineering, Kangwon National University

요약

본 논문에서는 영상으로부터 생성된 깊이맵을 합성곱 신경망(CNN)으로 재생성하는 방법을 제안한다. 합성곱 신경망은 영상인식, 영상분류에 좋은 성능을 보여주는데, 이 기술을 깊이맵 생성에 활용하여 기 제작된 깊이맵 생성 기법을 간단한 합성곱 신경망으로 구현하고자 한다. 성능 실험에서는 10개의 비디오 세트에 제안 방법을 적용한 결과, 만족스러운 결과를 얻었다.

1. 서론

영상으로부터 깊이맵을 생성하는 기법은 오랫동안 연구되어온 분야이다. 이 깊이맵은 장면내의 기하학적 관계를 이해할 수 있는 중요한 정보이다. 이 깊이정보는 물체 인식, 3D 모델링, 3D변환, 로봇 등에 활용될 수 있고, 잠재적으로는 객체들 간에 가려짐이 발생할시 객체 추론을 할 수 있다. 스테레오 센서, 깊이카메라 등으로 정확한 깊이를 얻을 수 있지만, 다른 분야로, 단안영상으로부터 영상의 깊이를 예측하는 연구도 꾸준히 진행되어 왔다. 다양한 방법들이 제안되었는데, 모션을 이용한 방법, 초점을 이용한 방법, 기하학적 특성을 이용한 방법 등이 있다.

기존 깊이맵 생성 방법들은 입력 영상과 기 제작된 GT(ground-truth) 깊이맵을 이용하여 다량의 데이터를 학습하고, 얻어진 합성곱 신경망(Convolutional Neural Network: CNN)으로 테스트하는 일반적인 신경망 기반이다 [1,2,3]. 이에 반해 본 연구는 기존의 깊이맵 생성 알고리즘을 CNN으로 대체하는 이론을 제안하고, 타당성을 검증한다. 기존 알고리즘은 [4]에서 제안한 방법을 사용한다.

2. 제안하는 합성곱 신경망

합성곱 신경망의 구성은 합성곱 계층, 풀링층으로 이루어진다. 합성곱층에서는 디지털 필터들을 모아놓은 계층으로서 합성곱 연산을 통해 입력 이미지를 변환하는 역할을 한다. 풀링층에서는 주위의 픽셀을 묶어서 하나의 대표 픽셀로 바꾼다. 즉, 이미지의 차원을 축소한다. 이는 다시 동일한 구조의 합성곱층과 풀링층에 의해 계산된다. 그림 1은 제안하는 CNN의 구조이다. 그림 2는 그림1의 CNN의 내부 구조이다.

합성곱 계층은 입력 이미지에서 특징을 살린 새로운 이미지를 만들어 낸다. 이 과정에서 입력 이미지를 다른 이미지로 변환하는 필터들을 사용하여 특징 맵을 만들어 내고, 이 특징 맵은 활성함수를 거쳐 최종 출력된다. 본 연구에서는 640x480의 이미지를 8x8의 이미지로 분할하였고, 3x3 크기의 필터를 사용하였으며 활성함수로는 Sigmoid 함수를 사용하였다.

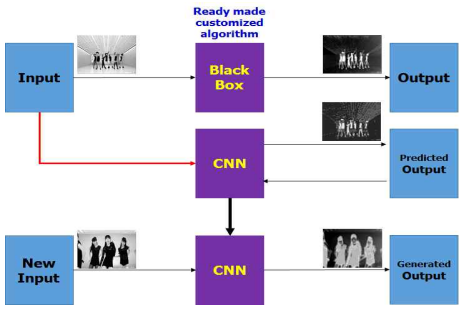

그림 1. 제안하는 CNN 기반 깊이맵 생성

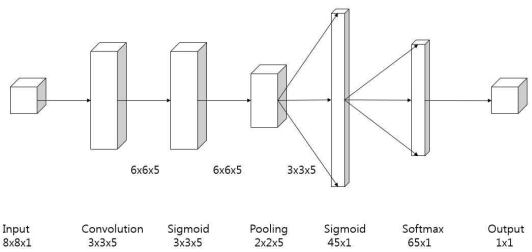

그림 2 그림 1의 CNN의 구조

3. 실험 결과

본 논문에서는 Gray-Scale 영상을 GT(Ground Truth) 깊이맵으로 변환하는 기존 알고리즘과 CNN을 통하여 예측된 값을 만들어 내는 실험을 하였다. Input Image는 36장의 이미지를 사용하였다. 전체 이미지중 학습 이미지와 테스트 이미지의 비율을 8:2로 나누었고, 이미지가 가질 수 있는 0~255의 값을 0~64로 양자화 하였으며 모든 이미지

는 8x8 사이즈로 분할하여 3x3 필터를 씌웠다. 실험 결과, 최적 학습율(α=0.01), 학습 횟수(epoch = 50)를 산정하였고, 이와 다른 이미지(17장, 39장, 20장)을 각각 테스트하여 결과를 도출하였다. customized algorithm으로 얻어진 깊이맵을 결과값과 비교하기 위하여 깊이값을 8x8의 크기로 분할한 후에 블록 평균 깊이를 이용하였다.

그림 3은 입력영상과 [1]을 이용하여 얻은 GT 깊이맵을 보여준다. Block depth map은 GT 깊이맵을 블록단위로 얻은 것이다. 마지막으로 Block predicted depth map은 CNN으로 생성된 깊이맵이다. 그림 4는 CNN의 epoch에 따라 얻어진 깊이맵을 보여준다. 전체적으로 epoch=50에서 주관적, 객관적으로 가장 우수한 깊이맵을 얻었기 때문에, 실험에서는 이 epoch값을 이용하였다.

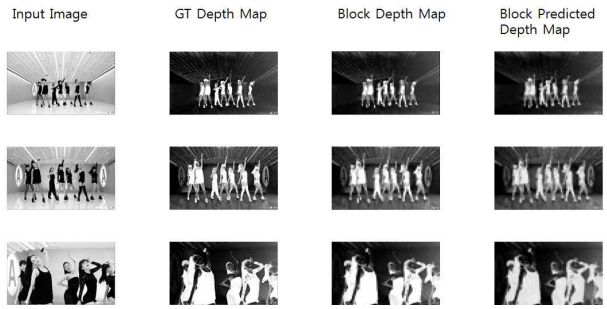

그림 3. 실험영상의 결과

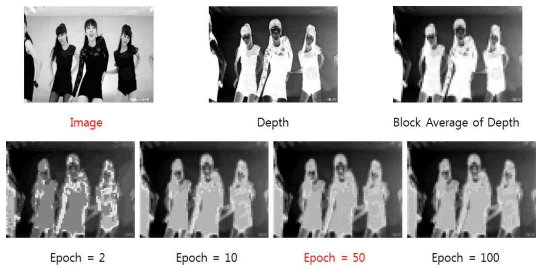

그림 4. epoch 횟수와 얻어진 예측 깊이맵

다음의 4가지 error metric을 이용하여 제안방법의 성능을 검증한다. D는 GT이고, $\hat{D}$는 예측된 깊이값이다.

1) Mean Absolute error (REL)

$$\mathrm{REL} = \frac{1}{|N|}\sum\frac{|D-\hat{D}|}{D}$$

2) Root Mean Square Eerror (RMS)

$$\mathrm{RMS} = \sqrt{\frac{1}{|N|}\sum(D-\hat{D})^2}$$

3) Average $\log_{10}$ error

$$\log 10 = \frac{1}{|N|}\sum|\log_{10}D - \log_{10}\hat{D}|$$

4) Accuracy with threshold: percentage (%) of $\hat{D}$

$$\delta = \max(\frac{D}{\hat{D}}, \frac{\hat{D}}{D}) < thr^{i_i}, \text{where } thr = 1.25,\ i = 1,2,3$$

표 1은 깊이맵의 성능결과를 보여주고, 표 2는 각 비디오 세트에서 전체 프레임, 학습 프레임, 테스트 프레임의 개수를 보여준다. 기 제작된 알고리즘을 8x8로 나눈 결과와 CNN을 통하여 나온 결과값의 accuracy는 평균 98.57%의 일치율을 보였고, 오류검출에 있어서도 REL은 평균 0.09, RMS는 2.55, Average $\log_{10}$ error 은 0.041의 낮은 수치를 보였다.

표 1. GT와 예측된 깊이의 성능 결과. α = 0.01, epoch = 50

Video	Accuracy with threshold:(%)	REL	RMS	Average $\log_{10}$ error
Rainbow	98.7500	0.0748	2.3700	0.0239
ani	99.8929	0.0470	1.6637	0.0194
bird	98.6670	0.0296	1.9601	0.0110
birdfall	98.3908	0.0758	2.2761	0.0216
boat	99.0440	0.0270	2.8310	0.0137
football	99.4633	0.0284	1.4362	0.0106
girl	97.3490	0.0754	4.5239	0.0328
horse	98.1815	0.1219	2.7199	0.0286
ski	99.5465	0.0574	2.2816	0.2376
visor	98.4266	0.1481	2.9962	0.0170
Average	8.5544	0.0685	2.5059	0.0416

표 2. 실험 비디오의 데이터. 학습프레임과 테스트프레임의 개수.

Video	No of frames	Training frames	Test frames
Rainbow	33	26	7
ani3	36	28	8
bird	60	48	12
birdfall	30	24	6
boat	300	240	60
football	132	105	27
girl	21	16	5
horse	71	56	15
ski	66	52	14
visor	22	16	6

4. 결론

본 논문에서는 CNN을 이용하여 기 제작된 깊이맵 알고리즘을 대체할 수 있는 방법을 제안하였다. 깊이맵의 성능은 본 연구의 목적이 아니고, 성능에 관계없이 어떠한 깊이맵 생성 알고리즘이라도 이 방법을 CNN으로 대체 가능하는지에 대한 타당성을 검증하는 것이 주목적이다. 높은 복잡도를 가지는 깊이맵 알고리즘을 낮은 복잡도를 가지는 CNN으로 대체할 수 있으면, 상당한 활용이 기대될 수 있다. 또한 제안 방법은 깊이생성이 아닌 타 기술에서도 적용하는 것이 가능하다.

감사의 글

2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3B03028806)

참고 문헌 (References)

[1] F. Liu, C. Shen, G. Lin, and I. Reid, “Learning Depth from Single Monocular Images Using Deep Convolutional Neural Fields”, IEEE Trans. Pattern Analysis and Machine Intelligence, Vol. 38, No. 10, Oct. 2016.
[2] D. Eigen, C. Puhrsch, and R. Fergus, “Depth map prediction from a single image using a multi-scale deep network”, 2014.
[3] Ahmed J. Afifi and Olaf Hellwich, “Object Depth Estimation from a Single Image using Fully Convolutional Neural Network“,
[4] 김만배, “관심맵과 에지 모델링을 이용한 2D 영상의 3D 변환”, 방송공학회논문지, 2015년 5월