

특집논문 (Special Paper)

방송공학회논문지 제23권 제1호, 2018년 1월 (JBE Vol. 23, No. 1, January 2018)

<https://doi.org/10.5909/JBE.2018.23.1.3>

ISSN 2287-9137 (Online) ISSN 1226-7953 (Print)

깊이맵 생성 알고리즘의 합성곱 신경망 구현

이 승 수^{a)}, 김 홍 진^{a)}, 김 만 배^{a)†}

Implementing a Depth Map Generation Algorithm by Convolutional Neural Network

Seungsoo Lee^{a)}, Hong Jin Kim^{a)}, and Manbae Kim^{a)†}

요 약

깊이맵은 현재 다양한 분야에서 활용되고 있다. 이러한 깊이맵을 인공 신경망으로 생성하는 연구가 최근 관심을 받고 있다. 본 논문에서는 기존의 기 제작된 깊이맵 생성 알고리즘을 합성곱 신경망으로 구현할 수 있는지에 대한 타당성을 검증한다. 먼저 깊이맵은 관심맵과 운동 히스토리 영상의 가중치 합으로 얻는다. 실험영상과 깊이맵을 합성곱 신경망의 입력과 출력으로 하여, 신경망을 학습시킨다. 정성적, 정량적 실험 결과는 제안한 합성곱 신경망이 깊이맵 생성 방법을 대체할 수 있다는 것을 보여준다.

Abstract

Depth map has been utilized in a variety of fields. Recently research on generating depth map by artificial neural network (ANN) has gained much interest. This paper validates the feasibility of implementing the ready-made depth map generation by convolutional neural network (CNN). First, for a given image, a depth map is generated by the weighted average of a saliency map as well as a motion history image. Then CNN network is trained by test images and depth maps. The objective and subjective experiments are performed on the CNN and showed that the CNN can replace the ready-made depth generation method.

Keywords : Depth map, CNN, Saliency map, Motion Hitsory Image, Ready-made depth map

a) 강원대학교 컴퓨터정보통신공학과(Department of Computer and Communications Eng., Kangwon National University)

† Corresponding Author : 김만배(Manbae Kim)

E-mail: manbae@kangwon.ac.kr

Tel: +82-33-250-6395

ORCID: <http://orcid.org/0000-0002-4702-8276>

※ 이 논문은 2017년도 강원대학교 대학회계 학술연구조성비로 연구하였음 (관리번호-520170077). 또한 2017년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (No. 2017R1D1A3 B03028806).

· Manuscript received October 31, 2017; Revised November 20, 2017; Accepted November 20, 2017.

1. 서 론

영상으로부터 깊이맵을 생성하는 기법은 오랫동안 연구되어온 분야이다. 이 깊이맵은 장면내의 기하적 관계를 이해할 수 있는 중요한 정보이다. 깊이정보는 물체인식, 3D모델링, 3D변환, 로봇 등에 활용될 수 있고, 잠재적으로는 객체들간에 가려짐 발생시 추론을 할 수 있다. 스테레오센서, 깊이카메라 등으로 정확한 깊이를 얻을 수 있지만, 다른 분

야로 단안영상으로부터 영상의 깊이를 예측하는 연구도 꾸준히 진행되어 왔다. 다양한 방법들이 제안되었는데, 모션을 이용한 방법, 초점을 이용한 방법, 기학적 특성을 이용한 방법등이 있다^[1-3].

최근에 기존의 *hand-crafted* 특징을 사용하지 않고, 합성곱 신경망(Convolutional Neural Network: CNN)을 이용하여 깊이맵을 생성하는 방법들이 소개되고 있다^[4,6]. CNN 기반 깊이맵 방법은 다른 신경망과 마찬가지로 학습 데이터를 활용한다. 입력영상과 실측 (*ground-truth*) 깊이맵을 학습하면서 실측 깊이맵과 유사한 깊이맵을 생성할 수 있도록 신경망을 제작한다. 이 과정은 다량의 깊이맵의 학습을 필요로 한다. 최근 이와 관련해서 다수의 연구가 제안하였다. Eigen 등은 *multi-scale CNN*을 제안하였는데^[4], 첫 단계에서는 *coarse depth*를 예측한다. 두 번째 단계에서는 *coarse depth*와 국부 깊이 예측과 결합하여 전체 영상의 예측 깊이값을 갱신하는 기법을 제안하였다. Afifi 등은 CNN을 이용하고 여기에 회귀모델을 사용하였다^[5]. Liu 등은 딥 CNN과 CRF(*conditional random fields*)를 이용하여 깊이맵을 생성하는데^[6], 전체영상을 *superpixel region*으로 분할한 후에 각 *superpixel*의 깊이를 예측한다.

상기 깊이맵 생성 방법들은 입력 영상과 기제작된 실측 깊이맵을 이용하여 다량의 데이터를 학습하고, 테스트하는 일반적인 신경망 기반이다. 전체적으로 우수한 성능을 얻

을 수 있지만, 특정 영상에 적용하게 되면 성능 저하가 발생한다. 신경망은 전체 데이터중 일부를 학습데이터로 학습한 후에, 나머지 데이터로 테스트하는 것이 주 방법이다. 따라서 새로운 데이터가 입력되면 과적합(*overfitting*)이 발생하게 되고 이것은 성능저하의 원인이 된다.

또한 정지영상의 깊이 예측에 한정되어 있어서, 모션의 처리가 어렵다. 이에 반해 본 연구는 기존의 깊이맵 생성 알고리즘을 CNN으로 대체하는 새로운 개념의 이론을 제안하고 타당성(*feasibility*)을 검증한다. 이를 위해서 먼저 특정 깊이맵 생성 알고리즘을 소개하고, CNN으로 학습을 하게 한 후에, CNN과 특정 알고리즘의 성능을 조사한다. 성능은 깊이값의 오차로 판단할 수 있다.

그림 1은 기존 깊이맵 생성 방법과 제안방법의 개념의 차이를 보여준다. 기존 깊이맵 생성방법은 그림 1(a)에서 보여지는데, 다량의 학습영상과 실측 깊이맵으로 CNN을 학습시킨 후에, 예측 깊이맵을 얻는 방식을 보여준다. 이와는 달리 제안방법은 그림 1(b)에서 보는 것처럼 기제작된 깊이맵 생성 방법으로 깊이맵을 생성한 후에, CNN으로 입력영상과 출력 깊이맵을 학습 및 테스트 하면서 CNN 모델을 완성한 후에, 이 CNN으로 기제작 깊이맵 생성을 대체하고자 하는 것이다.

깊이맵 생성 방법의 성능은 본 연구의 목적이 아니고, 성능에 관계없이 어떠한 깊이맵 생성 알고리즘이라도 CNN

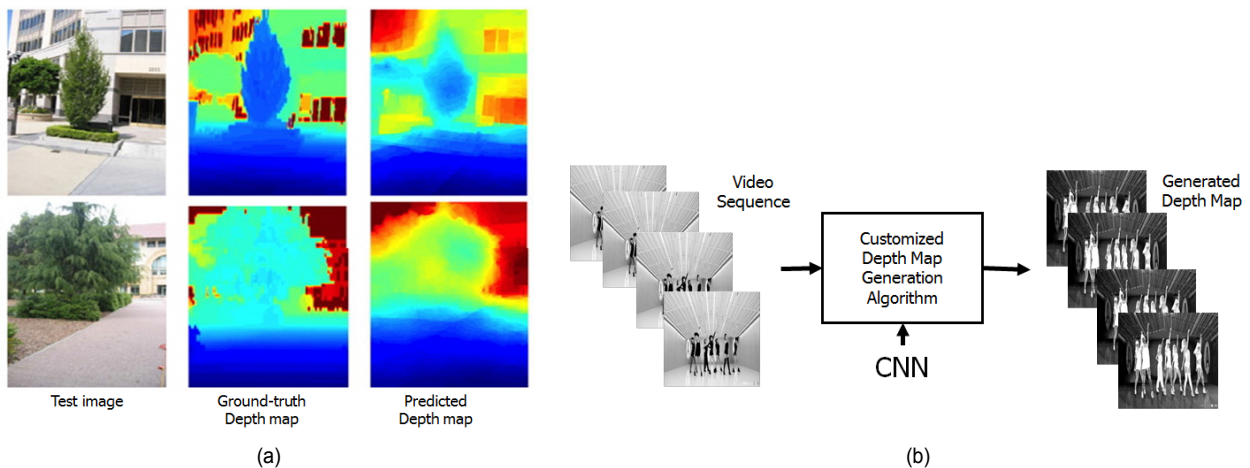


그림 1. 기존 깊이맵 생성과 제안 방법의 차이. (a) 일반적인 CNN기반 깊이맵 생성^[6] 및 (b) 기제작된 깊이맵 생성 알고리즘을 CNN으로 대체하는 구조도
Fig. 1. Difference of conventional depth map methods and the proposed method. (a) The approach of the former and (b) diagram replacing a customized depth map generation algorithm with CNN

으로 대체 가능하는지에 대한 타당성을 검증하는 것이 주 목적이다. 이를 위해 기존에 소개된 깊이맵 알고리즘을 활용하여 깊이맵을 생성하고, 이 깊이맵으로 CNN을 학습시킨다. CNN의 출력 깊이맵과 기존 알고리즘의 출력 깊이맵의 오차를 검증한다. 더욱이 높은 복잡도를 가지는 깊이 생성 기법들을 낮은 복잡도를 가지는 CNN으로 대체할 수 있으면, 상당한 활용이 기대될 수 있다. 결과적으로 어떠한 깊이맵 생성 기법을 연구하더라도 이를 대체할 수 있는 장점이 있다.

본 논문의 구성은 다음과 같다. 다음 장에서는 깊이맵 생성 알고리즘을 설명한다. III장에서는 제안하는 합성곱 신경망을 소개하고, 실험 결과는 IV장에서 소개한다. 마지막으로 V장에서는 결론 및 향후 연구를 정리한다.

II. 깊이맵 생성 방법

서론에서 언급하였듯이, 본 논문의 주 목적은 기존 깊이맵 생성 방법을 CNN으로 대체할 수 있는 지에 대한 타당성을 검증하는 것이다. 따라서 깊이맵 성능은 주 관심이 아니기 때문에 이 절에서는 Kim 등이 제안한 깊이맵 생성 방법의 일부를 사용한다^[7,8]. 이 방법은 관심맵(saliency map)으로 얻은 정적 깊이맵^[7]과 운동 히스토리 영상(motion history image)을 이용하여 얻은 모션 깊이맵^[8]의 가중치의 합으로 최종 깊이맵을 얻는다.

관심맵은 깊이맵 생성에서 활용해온 방법인데^[9,10], 영상을 인식하면서 눈에 띄는 영역이나 객체를 집중하는 방식을 수학적으로 모델링한 것이다. 입력 영상의 통계적 정보를 활용하여 픽셀 값의 빈도수(frequency)가 크거나, 영역이 큰 픽셀에서는 작은 관심도(saliency)를 할당하고, 반대로 인 경우에는 큰 값을 할당한다. 정지영상의 공간 관심맵을 구하는 방법 중의 하나인 Zhai가 제안한 전역 기반 방식은 색상 거리 등을 고려하여 관심도를 계산하는 방식이다^[11]. 픽셀 I_i 의 관심도 값은 입력 영상에서 다른 픽셀들과의 색상 차이(color difference)를 이용하여 계산한다. 픽셀 I_i 의 관심도 S_i 는 다음 식으로부터 얻어진다.

$$S_i = \sum_{j=1}^N D(I_i, I_j) = \sum_{j=1}^N |I_i - I_j| \quad (1)$$

여기서 N 은 영상 I 의 픽셀 개수이다. $D(I_i, I_j)$ 는 컬러 공간에서 두 픽셀의 색상 차이(color distance metric)이다. $I \in R, G, B$.

여기에서는 RGB의 변형을 사용한다. 색차신호 a, b, c 는 식 (2)을 사용하여 RGB로부터 얻는다.

$$\begin{aligned} a &= |R - \frac{G+B}{2}|, \quad b = |\frac{|B+G|}{2} - \frac{|B-G|}{2} - R|, \\ c &= |B - \frac{R+G}{2}| \end{aligned} \quad (2)$$

각 채널에서 관심맵은 다음과 같이 구해진다.

$$A = w_1 \cdot |a - \bar{a}|, \quad B = w_2 \cdot |b - \bar{b}|, \quad C = w_3 \cdot |c - \bar{c}| \quad (3)$$

여기서 w_i 는 가중치이고, \bar{x} 는 x 의 평균값이다.

관심도 S 는 A, B, C 의 최대값으로 구하고, 이 S 값을 정지 깊이맵으로 활용한다.

$$S = \text{MAX}(A, B, C) \quad (4)$$

S 는 $[0, 255]$ 의 값을 가진다.

운동 깊이맵은 운동 히스토리 영상(motion history image)을 이용하여 생성한다^[12]. 먼저 시간 t 에서의 현재영상 I_t 과 이전영상 I_{t-1} 으로부터 차영상 Δ_t 를 구한다.

$$\Delta_t^i = |I_t^i - I_{t-1}^i| \quad (5)$$

여기서 i 는 픽셀이다. Δ_t 를 구하면 깊이맵 D 는 식 (6)처럼 계산한다. 여기서 ζ 는 누적 파라메타이고, τ 는 감소 파라메타이다. D 의 초기값은 0이다.

$$D_t^i = \begin{cases} \min(D_{t-1}^i + \zeta, 255) & \text{if } \Delta_t^i > T \\ \max(D_{t-1}^i - \tau, 0) & \text{otherwise} \end{cases} \quad (6)$$

$\Delta_t > T$ 이면 D 는 ζ 만큼 증가하고, 반대이면 τ 만큼 감소하게 된다. 즉 픽셀의 D 값은 ζ 값이 크면 빠른 속도로 증가하고 τ 가 크면 급격히 감소하게 된다. 최근에 모션이 발생



그림 2. CNN 학습에서 사용하는 깊이맵. (a) 입력 영상 및 (b) 생성된 깊이맵

Fig. 2. Depth maps used in CNN training. (a) Input images and (b) generated depth maps

한 픽셀은 D 값이 증가한다.

식 (4)와 (6)에서 얻어진 두 개의 깊이맵의 가중치 합으로 최종 깊이맵을 생성한다. 그림 2는 상기 방법으로 생성한 깊이맵을 보여준다. 이 깊이맵은 CNN에서 실측(ground-truth) 깊이맵으로 활용된다.

III. 합성곱 신경망

합성곱 신경망(CNN)의 구성은 특징을 얻기 위한 합성곱 계층, 풀링(pooling)층 및 분류를 위한 완전결합(fully-connected) 신경망으로 이루어진다. 합성곱 계층에서는 디지털 필터들을 모아놓은 계층으로서 합성곱 연산을 통해 입력영상을 변환하는 역할을 한다. 입력영상에서 특징을 살린 특징맵(feature map)을 만들어 낸다. 이 과정에서 입력영상을 다른 영상으로 변환하는 필터들을 사용하여 특징맵을 만들어 내고, 이 특징은 활성화 함수(activation function)를 거쳐 출력된다. 풀링층에서는 주위의 픽셀을 묶어서 하나의 대표 픽셀로 바꾼다. 즉, 이미지의 차원을 축소한다. 최댓값 풀링은 픽셀의 값 중 최댓값을 갖는 값을 사용하여 영상의

크기를 감소시킨다. 평균값 풀링은 픽셀의 평균값을 계산하여 영상의 크기를 감소시킨다.

그림 3은 II장에서 소개한 깊이맵 생성 방법을 구현하기 위한 CNN의 전방향(feed-forward) 네트워크를 보여준다. 640x480 RGB영상을 8x8의 블록으로 분할하고, 64개의 픽셀 값의 평균이 블록의 밝기값이 된다. 이 값이 신경망의 입력 데이터가 된다.

[6]에서는 SLIC^[13]을 이용하여 영상을 임의의 형태를 가진 superpixel patch로 얻지만, 실제 깊이맵 처리는 superpixel의 중심점을 기반으로 블록 단위로 CNN에 입력한다. 본 연구에서는 superpixel을 사용하지 않고 직접 블록 단위로 처리한다. CNN으로 입력되면 합성곱, 시그모이드 활성화 함수, 및 풀링을 거친 후에 이 특징맵은 1개의 은닉층으로 구성된 완전결합 신경망으로 입력된다.

5개의 3x3 합성곱 필터를 적용하면 6x6 특징맵이 얻어지고, 2x2 풀링을 적용하면 5개의 3x3 특징맵이 얻어진다. 특징맵은 1차원 배열로 나열한 후에, 45개의 입력 노드로 보내진다. 활성화 함수(activation function)는 시그모이드(sigmoid)를 사용한다. 45개의 입력노드는 완전결합(fully-connected)층으로 입력된다. 은닉층의 노드는 10개이다.

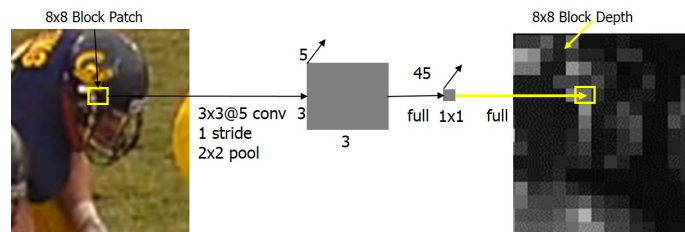


그림 3. 제안하는 합성곱 신경망 모델

Fig. 3. Proposed CNN model

생성된 깊이맵을 $[0, 255]$ 로 선형 매핑한 후에 양자화를 거쳐 $[0, 63]$ 의 깊이로 축소한다. 깊이맵에서 블록은 64개 픽셀의 깊이값의 평균값을 가진다. 픽셀이 아닌 블록단위로 깊이를 얻기 때문에 이 양자화는 필요하다. 출력노드는 깊이의 범위가 $[0, 63]$ 이므로 64개의 출력노드가 필요하다. 이 노드의 출력값은 소프트맥스(softmax)로 변환되어, 최대값을 가지는 노드가 해당 클래스가 된다.

CNN의 학습과정은 그림 4에서 보여진다. 발생하는 오차는 오류 역전파를 이용하여 합성곱 필터 및 은닉층의 가중치를 갱신하게 된다. 오류 역전파는 확률적 경사 하강법 (stochastic gradient descent)을 이용하고, 에폭(epoch)은 최대 100으로 하였다. 신경망의 가중치 초기화는 Xavier 초기화^[14]를 사용한다. 학습률 $\alpha=0.01$ 이고 모멘텀은 사용하지 않았다.

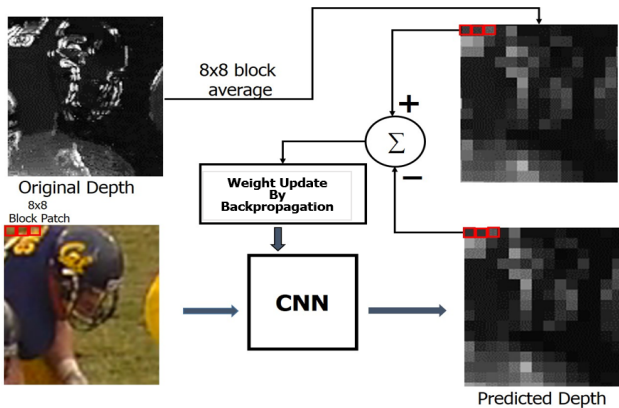


그림 4. 합성곱 신경망의 학습 과정
Fig. 4. Training procedure of CNN

IV. 실험 결과

실험에서는 II장의 깊이맵 생성 방법을 CNN으로 대체할 수 있는지에 대한 실험을 수행하였다. 제안 방법의 성능 개선을 검증하기 위해서 표 1의 10개의 데이터 세트를 사용하였다^[15-17]. 이중 *horse*, *ski*, *girl*은 [15]의 실험 동영상이고, *birdfall*은 [16], 그리고 *visor*은 [17]의 데이터이다. 모든 영상들은 640×480 으로 영상을 크기를 조절하였다. 먼저 깊이맵 생성 알고리즘을 얻은 3장의 연속 영상들의 깊이맵의

일부는 그림 5에서 보여진다. 표 1은 실험 비디오의 프레임 개수, 학습용 프레임 수 및 테스트 프레임 개수를 보여준다. 8×8 블록단위로 처리하기 때문에, 영상 한 개당 블록의 개수는 4,800이다. 일반적으로 총 학습데이터 중 학습과 테스트는 8:2로 하기 때문에 본 실험도 이 비율을 따랐다.

표 1. 실험 비디오의 총 프레임, 학습프레임 및 테스트 프레임의 개수
Table 1. No. of frames for each video including training and test frames

Video	No. of frames	Training frames	Test frames
Rainbow	33	26	7
ani	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

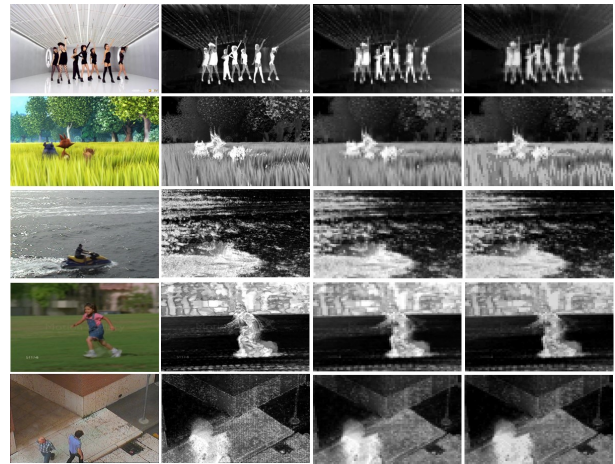


그림 5. 실험 영상과 II장의 깊이맵 생성으로 얻어진 연속 깊이맵
Fig. 5. Test images and subsequent depth maps made by the algorithm of Ch. 2

깊이맵 생성 방법의 결과를 실측(ground-truth) 깊이맵으로 하고, CNN을 통하여 예측된 깊이맵을 예측 깊이맵으로 한다. 정성적 실험과 정량적 실험을 통해서 제안한 CNN 모델을 검증한다. 그림 6은 정성적 실험 결과를 보여준다. 실험영상의 원 픽셀 깊이맵에서 8×8 블록 평균으로 얻은

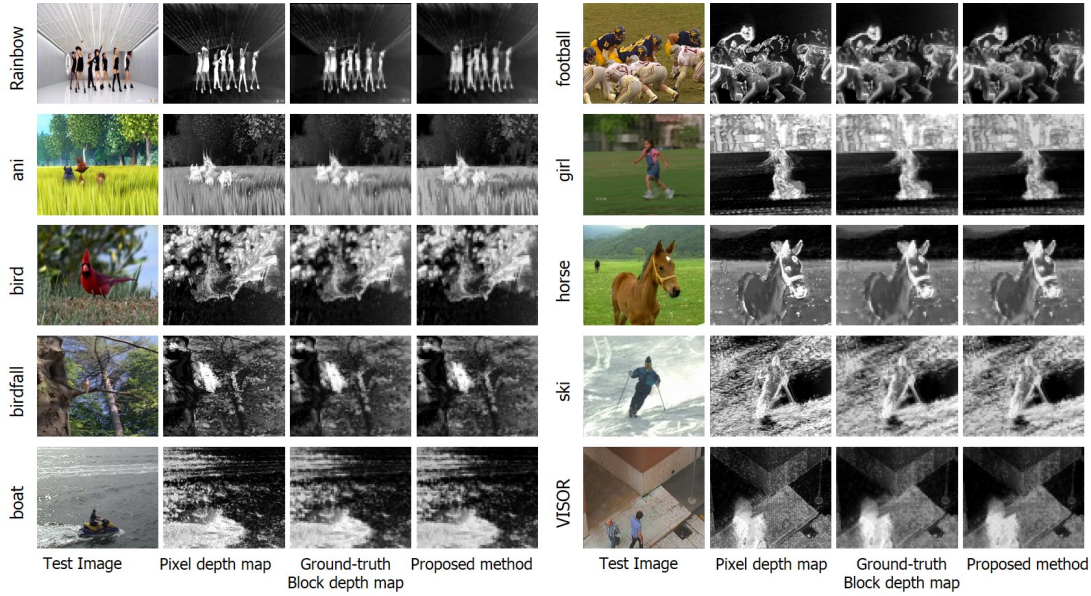


그림 6. 실험 영상으로 생성된 깊이맵의 비교. 제안 방법의 깊이맵과 실측 깊이맵과의 오차는 적음. 깊이값에 x4를 해서 [0, 255]로 조절하였음

Fig. 6. Test Images and depth maps. (a) input image, (b) pixel depth map, (c) block-average ground-truth depth map, and (d) predicted block depth map

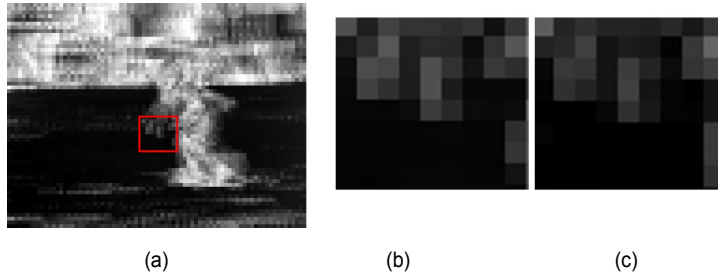


그림 7. *girl* 영상의 깊이맵의 확대 영상. 깊이값을 x4를 해서 [0, 255]로 조절하였음. a) 전체 깊이맵, (b) 확대된 실측 깊이, 및 (c) 예측된 깊이
Fig. 7. Close-up of a *girl* depth map. (a) full depth map and enlarged (b) ground-truth and (c) predicted depth maps

실측 블록 깊이맵과 제안 방법으로 얻은 깊이맵을 보면 주관적 평가 측면에서 차이가 없어 보인다. 이 결과는 제안한 CNN으로도 기존의 깊이맵 생성 알고리즘을 대체할 수 있다는 것을 보여준다.

그림 7은 그림 6의 *girl* 영상의 일부를 확대한 것이다. 그림 7(a)에서 적색 박스 영역에서 확대된 실측 및 예측 깊이맵은 그림 7(b)와 7(c)에서 보여준다. 값의 큰 차이는 관측되지 않는다.

제안한 CNN의 객관적 성능을 검증하기 위해서 깊이맵 성능 검증에서 자주 사용해온 다음 4가지의 성능 측정 방법을 이용한다.

1) average relative error (REL)

$$REL = \frac{1}{N} \sum \frac{|D - \hat{D}|}{D} \quad (7)$$

2) root mean squared error (RMS)

$$S = \sqrt{\frac{1}{N} \sum (D - \hat{D})^2} \quad (8)$$

3) average log10 error (log10)

$$\log_{10} = \frac{1}{N} \sum |\log_{10} D - \log_{10} \hat{D}| \quad (9)$$

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

$$\delta = \max\left(\frac{D}{\hat{D}}, \frac{\hat{D}}{D}\right) < \tau, \text{ where } \tau = 1.2 \quad (10)$$

여기서 D and \hat{D} 는 입력영상의 실측 및 예측 깊이값이다. N 은 모든 테스트 영상의 블록의 총 개수이다. 4) Accuracy에서 $\delta < \tau$ 이면 해당 블록의 깊이값은 만족하는 것으로, 이를 만족하는 블록의 개수를 총개수로 나누어서 %를 얻는다.

표 2. 실측과 예측 깊이의 성능 측정. $\alpha=0.01$, epoch=50
Table 2. Performance metric of ground-truth and predicted depths. $\alpha=0.01$, epoch=50

Video	Accuracy (%)	REL	RMS	log10
rainbow	98.7500	0.0748	2.3700	0.0239
ani3	97.8929	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	0.02376
visor	98.4266	0.1481	2.9962	0.0170
Average	98.5544	0.0685	2.5059	0.0416

표 2의 성능 측정 결과를 분석하면, Accuracy는 98.55%이다. REL은 0.0685, RMS는 2.5059, 및 log10은 0.0416이다. 제안 방법은 기존 깊이맵 생성 알고리즘을 신경망으로 대체하는 것이고, 대상으로 특정 방법을 이용하기 때문에 비교할 수 있는 연구가 부족하다. 따라서 참고적으로 [4]에 있는 성능 결과를 소개한다. Make3D Dataset에서 REL은 0.312, RMS=9.1, log10은 0.113을 얻었다. NYU v2 Dataset에 적용한 결과는 REL은 0.237, RMS는 0.822이고, log10은 0.082, Accuracy는 60.8%이다. 이 방법은 픽셀단위이고, 깊이의 범위는 [0, 255]이다. 본 연구에서는 [0, 63]의 깊이를 이용하는데, 이 차이를 감안해도 제안방법의 결과는 상당한 우수성을 보여준다.

V. 결론 및 향후 연구

본 논문에서는 기존의 깊이맵 생성 방법을 합성곱 신경망으로 구현할 수 있는지의 타당성을 검증하였다. 깊이맵 생성 방법으로 얻어진 깊이맵과 입력영상을 학습 데이터로 활용하였고, 10개의 실험 비디오로부터 얻은 정성적, 정량적 결과는 타당성을 뒷받침한다. 연구의 주 목적은 깊이맵 생성 알고리즘을 CNN으로 대체하는 것이다. 이 관점에서 제안 방법은 다른 분야에서도 활용할 수 있는 기술이다. 향후 블록기반을 픽셀단위로 처리하여 개선된 CNN 모델을 구현할 계획이다. 또한 실험에서는 특정 깊이맵 생성 방법을 통하여 CNN을 학습하였는데, 다른 깊이맵 생성방법들의 성능 검증을 위한 실험이 필요하다.

참고 문헌 (References)

- [1] S. Kim and J. Yoo, "3D conversion of 2D video using depth layer partition," Journal of Broadcast Engineering, Vol. 15, No. 2, Jan. 2011.
- [2] J. Jung, J. Lee, I. Shin, J. Moon and Y. Ho, "Improved depth perception of single view images", ECTI Transactions on Electrical Engineering, Electronics and Communications, Vol. 8, No. 2, Aug. 2010.
- [3] W. Tam and L. Zhang, "3D-TV Content Generation: 2D-To-3D Conversion," Proc. of IEEE ICME, 2006.
- [4] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network", Advances in Neural Information Processing Systems, 27, 2014.
- [5] A. Afifi and O. Hellwich, "Object Depth Estimation from a Single Image using Fully Convolutional Neural Network", Int' Conf. on Digital Image Computing: Techniques and Applications (DICTA), Nov. 2016.
- [6] 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.
- [7] M. Kim, "Generation of Stereoscopic image from 2D Image based on saliency and edge modeling", Journal of Broadcast Engineering, Vol. 20, No. 3, May 2015.
- [8] W. Kim, J. Gil and M. Kim, "Motion depth generation using MHI for 3D video conversion", Journal of Broadcast Engineering, Vol. 22, No. 4, July 2017.
- [9] Y. Zhang, G. Jiang, M. Yu, and K. Chen, "Stereoscopic visual attention model for 3D video", Advances in Multimedia Modeling, 2010.
- [10] J. Kim, A. Baik, Y. Jung and D. Park, "2D-to-3D image/video conversion by using visual attention analysis," Int' Conf. on Image Processing, 2009.

- [11] Y. Zhai, and M. Shah, "Visual attention detection in video sequences using spatiotemporal cues," 14th Annual ACM Int' Conf. on Multimedia, pp. 815-824, 2006.
- [12] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," IEEE Trans. Pattern Recognition and Pattern Analysis, Vol 23, No. 3 Mar. 2001.
- [13] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC superpixels compared to state-of-the-art superpixel methods", IEEE Trans. Pattern Analysis and Machine Intelligence, 34, (11), pp. 2274 - 2281, 2012.
- [14] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep forward neural networks", Int' Conf. Artificial Intelligence and Statistics, Society for Artificial Intelligence and Statistics, 2010.
- [15] K. Fukuchi, K. Miyazato, A. Kimura, S. Takagi, and J. Yamato, "Saliency-based video segmentation with graph cuts and sequentially updated priors," in Proc. IEEE Int. Conf. Multimedia Expo, pp. 638 - 641, June - July, 2009.
- [16] D. Tsai, M. Flagg, and J. M. Rehg, "Motion coherent tracking with multi-label MRF optimization," Proc. Brit. Mach. Vis. Conf., 2010.
- [17] D. Baltieri, R. Vezzani and R. Cucchiara, "3DPes: 3D People Dataset for Surveillance and Forensics," in Proceedings of the 1st International ACM Workshop on Multimedia access to 3D Human Objects, pp. 59-64, Nov-Dec, 2011. (<http://imagelab.ing.unimore.it/visor/3dpes.asp>)

저 자 소 개

이 승 수



- 2017년 8월 : 강원대학교 IT대학 컴퓨터정보통신공학과 학사
- 2017년 9월 ~ 현재 : 강원대학교 컴퓨터정보통신공학과 석사과정
- 주관심분야 : 점유센서, 딥러닝, 머신러닝

김 흥 진



- 2018년 2월 : 강원대학교 IT대학 컴퓨터정보통신공학과 학사
- 주관심분야 : 깊이맵 생성 기법, 딥러닝, CNN

김 만 배



- 1983년 : 한양대학교 전자공학과 학사
- 1986년 : University of Washington, Seattle 전기공학과 공학석사
- 1992년 : University of Washington, Seattle 전기공학과 공학박사
- 1992년 ~ 1998년 : 삼성종합기술원 수석연구원
- 1998년 ~ 현재 : 강원대학교 IT대학 컴퓨터정보통신공학과 교수
- 2016년 ~ 현재 : 강원대학교 정보통신연구소 소장
- ORCID : <http://orcid.org/0000-0002-4702-8276>
- 주관심분야 : 3D영상처리, 비전점유센서, 객체트래킹