이미지로부터 피사계 심도 영역을 효율적으로 추출하기 위한 합성곱 신경망 기법

김동희[O], 김종현[*]
[O]강남대학교 소프트웨어응용학부,
[*]강남대학교 소프트웨어응용학부
e-mail: jonghyunkim@kangnam.ac.kr[*]

Convolutional Neural Network Technique for Efficiently Extracting Depth of Field from Images

Donghui Kim[O], Jong-Hyun Kim[*]
[O]Dept. of Software Application, Kangnam University,
[*]Dept. of Software Application, Kangnam University

● 요 약 ●

본 논문에서는 카메라의 포커싱과 아웃포커싱에 의해 이미지에서 뿌옇게 표현되는 DoF(Depth of field, 피사계 심도) 영역을 합성곱 신경망을 통해 찾는 방법을 제안한다. 우리의 접근 방식은 RGB채널기반의 상호-상관 필터를 이용하여 DoF영역을 이미지로부터 효율적으로 분류하고, 합성곱 신경망 네트워크에 학습하기 위한 데이터를 구축하며, 이렇게 얻어진 데이터를 이용하여 이미지-DoF가중치 맵 데이터 쌍을 설정한다. 학습할 때 사용되는 데이터는 이미지와 상호-상관 필터 기반으로 추출된 DoF 가중치 맵을 이용하며, 네트워크 학습 단계에서 수렴률을 높이기 위해 스무딩을 과정을 한번 더 적용한 결과를 사용한다. 본 논문에서 제안하는 합성곱 신경망은 이미지로부터 포커싱과 아웃포커싱된 DoF영역을 자동으로 추출하는 과정을 학습시키기 위해 사용된다. 테스트 결과로 얻은 DoF 가중치 이미지는 입력 이미지에서 DoF영역을 빠른 시간 내에 찾아내며, 제안하는 방법은 DoF영역을 사용자의 ROI(Region of interest)로 활용하여 NPR렌더링, 객체 검출 등 다양한 곳에 활용이 가능하다.

키워드: 합성곱 신경망(Convolutional neural network), 피사계 심도(Depth of field), 이미지 프로세싱(Image processing), 인공 신경망(Artificial neural networks)

I. Introduction

이미지나 동영상 데이터를 활용하여 영상 내 콘텐츠의 분석, 객체의 검출, 사용자의 ROI 계산 등을 하는 과정에서 이미지의 피사계 심도는 중요한 특징 중 하나이다[1,2]. 원본 이미지에서 DoF가 포함된 영역을 뚜렷하게 식별하기란 어려운 일이며, 이는 DoF가 포함된 이미지에서 정확한 포커싱 영역을 추출하는데 있어서, 예측할 수 있는 정보가 부족하기 때문이다. 이미지에 내포되어 있는 장면으로부터 기하학적 관계를 분석하고자 깊이(Depth)를 추정하는 연구들은 꾸준히 연구되고 있으며[3,4], 이 기법은 포즈(Pose) 결정, 객체 인식 등에서 활용되고 있을 뿐, 이미지의 ROI라고 결정하기에는 다소 무리가 있다. 실제로 DoF는 콘텐츠의 깊이와는 다르게 사용자가 집중적으로 보고 싶은 부분을 두드러지게 하는 기법이기 때문에 이러한 연구들은 본 연구에서 제안하고자 하는 DoF와는 차이가 있다.

이미지에서 포커싱과 아웃포커싱의 차이는 선명하거나 뿌옇게 표현되는 차이가 있으며 이러한 차이는 공간상에서 미분을 했을 때 더 큰 차이를 보이게 된다. 우리는 이러한 특징을 이용하여 합성곱 신경망 기반의 새로운 네트워크 모델을 제안하며, 이러한 기법을 계산하기 위한 본 연구의 기여도는 아래와 같다 :

- 한 장의 이미지로부터 DoF영역을 추출할 수 있는 기술 개발
- 입력 이미지에서 DoF를 추출하기 위해, MSE(Mean squared error)의 수렴률을 향상시킬 수 있는 합성곱 신경망 기술 개발

II. The Proposed Scheme

1. Extraction of DoF region from a single image

우리는 이미지로부터 DoF영역을 계산하기 위해 상호-상관 필터 G를 이용한다. 이 필터는 두 개의 연속된 데이터들이 얼마나 연관되어 있는지를 계산하는 방법으로 영상처리, 컴퓨터 비전 등 다양한 분야에서 활용되고 있는 방법이다[5]. 이것은 서로의 색상 값이 얼마나 유사하게 관련되어 있는지를 측정하는 방법이며 수식 1과 같이 정의된다.

$$G(x,y)=H\otimes F=\sum_{u=-1}^{1}\sum_{v=-1}^{1}H(u,v)F(x+u,y+v) \quad (1)$$

여기서 H는 각 인접 픽셀들의 가중치로 마스크(Mask)라고 부르며, F는 인접 픽셀들의 색상 값이다. 이 마스크는 적용 분야에 따라 다양하게 모델링되는데, 우리는 가우시안 필터링(Gaussian filtering) 기법을 이용한다 (수식 2 참조).

$$H(x,y)=\frac{1}{16}\begin{bmatrix}1&2&1\\2&4&2\\1&2&1\end{bmatrix}\approx\frac{1}{2\pi\sigma^2}e^{-\frac{u^2+v^2}{\sigma^2}} \quad (2)$$

이미지에 존재하는 DoF영역을 추출하기 위해 본 논문에서는 몇 가지 가정을 한다 :

1) DoF영역에서는 피사계 심도를 얕게 하여 초점이 맞은 피사체를 제외한 배경은 흐려지게끔 뭉개지는 효과가 나타난다.
2) 아웃포커싱에 의해 뿌옇게 흐려진 영역과 선명한 영역의 차이는 명확하다.

위에서 언급한 특징을 이용하여, 원본이미지 I^{rgb_0}와 원본 이미지로부터 생성한 스무딩(Smoothing) 이미지 I_1^{rgb}의 RGB색상 채널의 차이를 기반으로 DoF가중치 맵 D를 다음과 같이 계산한다 (수식 3 참조).

$$D(x,y)=\frac{\sum_{u=-m}^{m}\sum_{v=-m}^{m}H_m(u,v)O(x+u,y+v)}{m^2} \quad (3)$$

$$O(x,y)=\parallel I_0^{rgb}(x,y)-I_1^{rgb}(x,y)\parallel \quad (4)$$

여기서 H_m과 m은 각각 마스크와 마스크의 크기를 나타내며, 본 논문에서 마스크의 크기는 15로 설정했다. O는 앞에서 언급한 뿌연 이미지와 선명한 이미지 사이의 색상 차이를 계산한 수식이다. I^{rgb_0}는 원본 이미지로부터 얻은 RGB색상이고, I_1^{rgb}는 뿌연 이미지로부터 얻은 RGB색상이며, 스무딩 필터는 가우시안 스무딩 기법을 사용했다. 수식에서 보듯이 DoF영역의 가중치는 RGB채널 값을 3차원 벡터로 표현하여 그 벡터의 크기를 통해 계산한다.

(a) input image (b) D

Fig. 1. DoF weight map D calculated using the DoF region (white : focusing, black : defocusing).

Fig. 1는 입력 이미지를 통해 얻은 DoF가중치 맵이다. DoF에 의해 Fig. 1a에서 아웃포커싱된 뿌연 영역의 가중치가 잘 표현되었고, 그 외 다양한 이미지들에서도 안정적으로 DoF맵을 추출했다 (Fig. 2 참조).

(a) (b)

Fig. 2. DoF weight maps calculated from various input images (inset image : input image).

2. Convolutional Neural Networks with DoF Weight map

우리는 앞에서 설명한 방법을 이용하여 RGB채널을 가진 입력 이미지인 $\{\delta^1,\delta^2,...\}$와 DoF가중치 맵 이미지인 $\{D^1,D^2,...\}$를 생성한다. 각 이미지는 학습 네트워크에 넣기 전에 패치단위로 분할한다. 학습 데이터가 주어지면, 우리의 목표는 예측된 값인 δ_s와 실제 값인 D사이의 오차를 최소화하는 매핑 함수 $f(x)$를 찾는 것이다. 이 과정을 수행하기 위한 목적 함수(Objective function)는 예측된 이미지와 실제 이미지 사이의 MSE이다. 우리의 목표는 $\delta_s=f(x)$값을 예측하는 모델 f를 학습하는 것이며, 결과적으로 학습 데이터에 대한 MSE인 $\frac{1}{2}\parallel D-f(x)\parallel^2$를 최소화하는 것이다. 이 방정식을 최적화시키기 위해 우리는 아래와 같은 SRCNN[6]기반의 네트워크 방식을 사용하였다 (Fig. 3 참조).

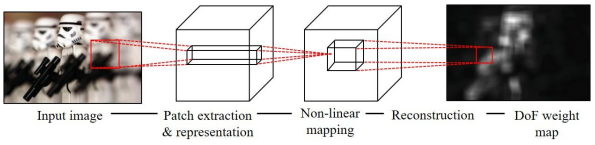


Fig. 3. SRCNN based network architecture.

Fig. 3에서 보여준 네트워크 방식을 적용하기 위해 입력 이미지의 DoF 가중치 맵에서 DoF 가중치 맵을 스무딩한 이미지로 변환하기 위한 가중치 매개변수를 계산해야 한다 : δ : $\rightarrow D$:

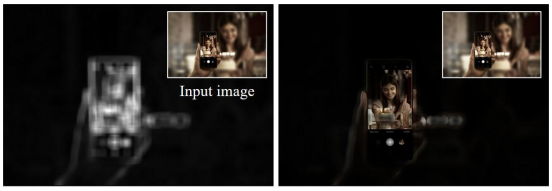
-(δ^* : $\rightarrow D^*$: ,θ). 여기서 δ, D는 입력 이미지와 그에 해당하는 DoF가중치 맵이며, δ^*, D^*는 δ를 스무딩한 입력 이미지이며, 그에 해당하는 DoF가중치 맵이고, θ는 우리가 찾고자하는 가중치 매개변수이다. 스무딩한 이유는 앞에서도 언급했듯이, 이미지에서 포커싱과 아웃포커싱의 차이는 선명하거나 뿌옇게 표현되는 차이가 있으며, 이 차이는 공간상에서 미분을 했을 때 더 큰 차이를 보이기 때문이다. 이러한 방식을 기반으로 학습을 시켰을 때 안타깝게도 우리가 기대한 결과를 얻을 수 없었으며, 학습과정에서 반복횟수가 증가해도수렴하지 않은 결과를 보였다 (Fig. 4 참조).

Fig. 4. Training results according to the number of iteration.

(a) Input image (b) Training result

Fig. 5. Training result obtained through 15,000 iterations.

우리는 15,000번 정도 학습시켰을 때 DoF영역이 아닌, 마치 흑백 이미지와 같은 결과로 수렴하는 학습결과를 얻었다 (Fig. 5 참조). 이 같은 이유는 매핑 함수 $f(x)$를 학습하는 과정에서 δ와 D의 차이가 서로 너무 다르기에 원하는 방향으로 수렴하지 않는 결과가 나타났으며, 다양한 이미지에서 실험을 해봤지만 여전히 같은 결과를 얻었다. 이 같은 문제를 해결하기 위해 본 논문에서는 가중치 맵인 D에 원본 이미지를 곱한 이미지인 $D' = D \times \delta$를 이용하여 네트워크의 학습 데이터로 사용하였고 (Fig. 6 참조), 본 논문에서 사용한 좀 더 자세한 네트워크 모델은 Fig. 7과 같다.



(a) D (b) D'

Fig. 6. Advanced weight image of DoF region.

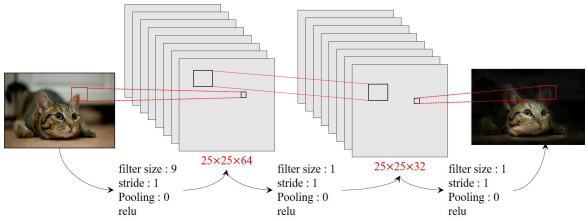


Fig. 7. The proposed network architecture.

Fig. 8은 본 논문에서 제안한 방법을 이용하여 학습되는 중간 결과를 보여주는 그림이다. Fig. 4,5와 비교해봤을 때, 이 그림은 15,000번의 반복적인 학습을 수행했음에도 불구하고 DoF와는 관계없는 흑백이미지로 수렴했지만, 제안한 방법은 훨씬 적은 반복횟수인 1,400번만에 DoF 영역을 빠르게 찾아냈다. 이 같은 결과의 의미는 DoF영역 추출과 별개로 수렴률이 굉장히 빠르다는 것을 나타낸다.

Fig. 8. Training results according to the number of iteration with our method.

또한, 학습하는 과정에서 425개의 적은 데이터셋의 양으로 좋은 결과를 만들어 냈다. 일반적으로 수천, 수만장의 데이터를 이용하는데 비해서 본 연구에서는 수렴율을 빠르게 하기 위한 전처리 방법으로 인해 적은 데이터만으로 높은 학습률을 나타냈다.

III. Conclusions

본 연구에서는 상호-상관 필터를 활용하여 DoF영역을 찾는 알고리즘을 제안했고, 이 기술을 인공지능을 통해 자동으로 학습하고 테스트할 수 있는 새로운 합성곱 신경망 기법을 제안했다.

Fig. 9는 본 논문에서 제안한 방법을 통해 얻어진 테스트 결과이며, 다양한 이미지들에서도 안정적으로 DoF가중치 맵을 추출했다. 레인보우 컬러 결과에서 빨간색에 가까울수록 포커싱 영역을 의미하며, 파란색에 가까울수록 아웃포커싱된 영역을 나타낸다. 향후 우리는 사용자의 시점 변화에 따른 ROI를 DoF기반으로 추출하고, 추출된 영역에서의 콘텐츠를 해석할 수 있는 기법으로 추가 연구할 계획이다.

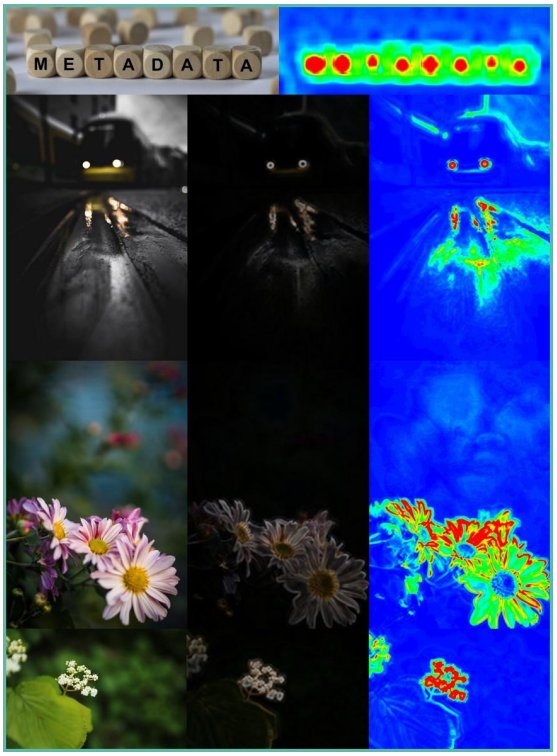


Fig. 9. Test results with our framework (rainbow color : DoF weight map).

REFERENCES

[1] Rafiee, Gholamreza and Dlay, Satnam Singh and Woo, Wai Lok, Region-of-interest extraction in low depth of field images using ensemble clustering and difference of Gaussian approaches. Pattern Recognition, pp. 2685-2699, 2013.

[2] Park, Jungwoo and Kim, Changick, Extracting focused object from low depth-of-field image sequences. Visual Communications and Image Processing, pp. 60771O, 2006.

[3] Simonyan, Karen and Zisserman, Andrew, Data-driven estimation of cloth simulation models. arXiv preprint arXiv:1409.1556, 2014.

[4] Hadsell, Raia and Sermanet, Pierre and Ben, Jan and Erkan, Ayse and Scoffier, Marco and Kavukcuoglu, Koray and Muller, Urs and LeCun, Yann, Learning long-range vision for autonomous off-road driving. Journal of Field Robotics, pp. 120-144, 2009.

[5] Danelljan, Martin and Robinson, Andreas and Khan, Fahad Shahbaz and Felsberg, Michael, Beyond correlation filters: Learning continuous convolution operators for visual tracking. European conference on computer vision, pp. 472-488, 2016.

[6] Dong, Chao and Loy, Chen Change and He, Kaiming and Tang, Xiaoou, Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, pp. 295-307, 2015.