이미지의 피사계 심도를 빠르게 계산하기 위한 쿼드트리 기반의 합성곱 신경망 최적화

김동희O, 김수균**, 김종현*
O강남대학교 소프트웨어응용학부,
*강남대학교 소프트웨어응용학부,
**제주대학교 컴퓨터공학과
e-mail: jonghyunkim@kangnam.ac.kr

Quadtree-based Convolutional Neural Network Optimization to Quickly Calculate the Depth of Field of an Image

Donghui KimO, Soo-Kyun Kim**, Jong-Hyun Kim*
OSchool of Software Application, Kangnam University,
*School of Software Application, Kangnam University,
**Dept. of Computer Engineering, Jeju National University

● 요 약 ●

본 논문에서는 카메라의 포커싱과 아웃포커싱에 의해 이미지에서 뿌옇게 표현되는 DoF(Depth of field, 피사계 심도) 영역을 쿼드트리(Quadtree) 기반의 합성곱 신경망을 통해 빠르게 찾는 방법을 제안한다. 우리의 접근 방식은 RGB채널기반의 상호-상관 필터를 이용하여 DoF영역을 이미지로부터 효율적으로 분류하고, 적응형 트리인 쿼드트리를 기반으로 유의미한 영역만을 분류한다. 이 과정에서 손실 없이 온전하게 DoF영역을 추출하기 위한 필터링 과정을 거친다. 이러한 과정에서 얻어진 이미지 패치들은 전체 이미지에 비해 적은 영역으로 나타나며, 이 적은 개수의 패치들을 이용하여 네트워크 단계에서 사용할 이미지-DoF가중치 맵 데이터 쌍을 설정한다. 네트워크 과정에서 학습할 때 사용되는 데이터는 이미지와 상호-상관 필터 기반으로 추출된 DoF 가중치 맵을 이용한다. 본 논문에서 제안하는 쿼드트리 기반 합성곱 신경망은 이미지로부터 포커싱과 아웃포커싱된 DoF영역을 자동으로 추출하는 과정을 학습시키기 위해 사용된다. 결과적으로 학습에 필요한 데이터 영역이 줄어듦으로써 학습 시간과 메모리를 절약했으며, 테스트 결과로 얻은 DoF 가중치 이미지는 입력 이미지에서 DoF영역을 더욱더 빠른 시간 내에 찾아낸다.

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

I. Introduction

합성곱 인공 신경망은 객체/문자 인식, 이미지 합성, 스타일 전달 등 다양한 분야에서 활용되고 있는 방법이다[1]. 하지만, 컨볼루션(Convolution)은 계산양이 크기 때문에 사실상 높은 계산 리소스를 요구하는 방식이다. 본 논문에서는 이미지에 내포되어 있는 DoF영역을 빠르고 효율적으로 찾기 위해 쿼드트리를 이용하였으며, 이 과정에서 학습에 필요한 데이터의 개수와 영역을 줄인다. DoF영역에 존재하는 픽셀 데이터만을 이용하여 네트워크 학습과정을 처리하기 때문에 전체적인 프로세스의 효율성을 한층 개선시킨다. 본 논문에서는 이러한 특징을 이용하여 합성곱 신경망 기반의 새로운 네트워크 모델을 제안하며, 이러한 기법을 계산하기 위한 본 연구의 기여도는 아래와 같다 :

- 영상에 내포되어 있는 DoF영역을 이미지 손실 없이 이진화하여 안정적으로 찾을 수 있는 기술 개발
- 이진화된 DoF영역에서 쿼드트리를 구축하고 패치들을 얻어낼 수 있는 기술 개발
- 패치들을 이용하여 DoF 합성곱 인공 신경망을 학습하고 테스트할 수 있는 기술 개발

II. The Proposed Scheme

1. Detection of DoF region in a single image

한 장의 이미지로부터 DoF영역을 계산하기 위해 상호-상관 필터 G를 이용한다. 이 필터는 두 개의 이미지들이 얼마나 연관되어 있는지를 계산하는 방법으로 영상처리, 컴퓨터 비전 등 다양한 분야에서 활용되고 있는 방법이다. 이것은 서로의 색상 값이 얼마나 유사한지를 측정하는 방법이다 (수식 1 참조).

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

여기서 H는 각 인접 픽셀들의 가중치로 마스크라고 부르며, F는 인접 픽셀들의 색상이다. 이 마스크는 적용 분야에 따라 다양하게 모델링되는데, 우리는 가우시안 필터링을 이용한다 (수식 2 참조).

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

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

1) 이미지의 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}_0와 원본 이미지로부터 생성한 스무딩 이미지 I^{rgb}_1의 RGB색상 채널의 차이를 기반으로 DoF가중치 맵 D를 다음과 같이 계산한다 (수식 3 참조).

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

$$O(x,y)=\| I^{rgb}_0(x,y)-I^{rgb}_1(x,y) \| \quad (4)$$

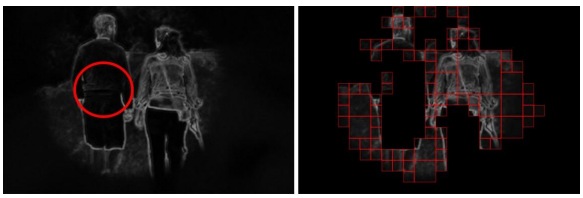
여기서 H_m과 m은 각각 마스크와 마스크의 크기를 나타내며, 본 논문에서 마스크의 크기는 15로 설정했다. O는 앞에서 언급한 뿌연 이미지와 선명한 이미지 사이의 색상 차이를 계산한 수식이다. 수식에서 보듯이 DoF영역의 가중치는 RGB채널 값을 3차원 벡터로 표현하여 그 벡터의 크기를 통해 계산한다. Fig. 1은 입력 이미지를 통해 얻은 DoF가중치 맵이다. DoF에 의해 Fig. 1a에서 포커싱된 뿌연 영역의 가중치가 잘 표현되었다.

(a) Input image (b) D

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

2. Collecting sparse data set from quadtree based DoF patches

우리는 앞에서 계산한 D를 이용하여 쿼드트리 기반 DoF 패치들을 구성하기 위해 전처리 과정을 거쳐야 한다. 이 과정은 온전하게 DoF 영역을 찾기 위함이며, 이 필터링 과정을 거치지 않으면 아래 Fig. 2에서 보듯이 빈 공간에 쿼드트리가 생성되지 않는 문제가 발생한다.



(a) Empty region (b) Incorrect quadtree segmentation

Fig. 2. Incorrectly subdivided quadtree due to empty space (red circle in (a) : empty region).

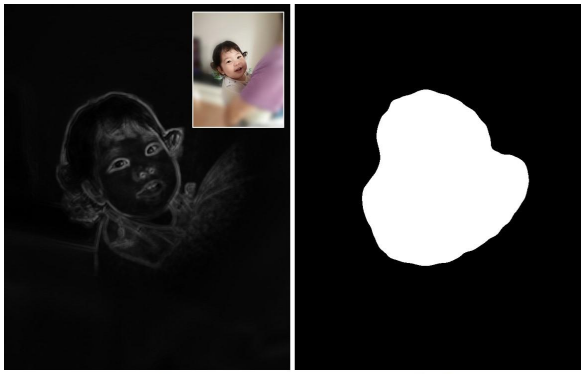
이 문제를 완화하기 위해 D에 대해서 감마 보정(Gamma correction)을 거친다. 이 과정은 포커싱과 아웃포커싱 영역의 차이를 더욱 크게 두기 위해서이며, 아래와 같은 수식을 통해 계산한다 (수식 5 참조).

$$f_{gm}=M\left(\frac{x}{M}\right)^g \quad (5)$$

여기서 f_{gm}와 x는 출력과 입력 이미지이며, M은 최댓값을 의미하고, 본 논문에서는 255를 사용하였다. 이 수식에서 g가 1이면 지수는 1이 되므로 선형적으로 밝기가 변화되며, 본 논문에서는 0.85로 사용하였다. 이렇게 얻어지는 f_{gm}은 색상은 좀 더 선명하지만, 경계나 에지부분에서 블러되는 문제가 있기 때문에, 이 문제를 완화하기 위해 라플라시안(Laplacian) 커널을 이용한 영상 샤프닝(Sharpening)를 적용한다 (수식 6 참조).

$$f_{sp}=f+\alpha(f-f_{blur}) \quad (6)$$

여기서 α는 마스크로 표현 될 수 있으며, 본 논문에서는 가우시안 형태의 마스크를 사용했다. 마지막으로 영상 이진화를 통해 DoF 가중치 맵인 D를 닫힌 영역으로 필터링 한다 (Fig. 3 참조).



(a) D (b) D^*

Fig. 3. D^* filtered from D ((a) : D, inset image in (a) : input image, (b) : D^*).

퀴드트리를 계산하기 위해 D^*를 이용하며, Fig. 3a에서 보듯이 흰색이 DoF 영역이기 때문에 이를 기준으로 퀴드트리를 분할한다 (Fig. 4 참조).

Fig. 4. Quadtree construction with D^*.

퀴드트리를 분할할 때 사용하는 이미지는 D^*이지만, 이것은 실제 픽셀 정보를 얻기 위함이 아닌, 공간상에서 유의미한 정보를 추출하려고 사용되는 이미지이다. 결과적으로는 퀴드트리의 말단 노드에 해당하는 이미지들의 패치들인 D_p를 이용하여 네트워크 과정에서 학습과 테스트를 진행한다.

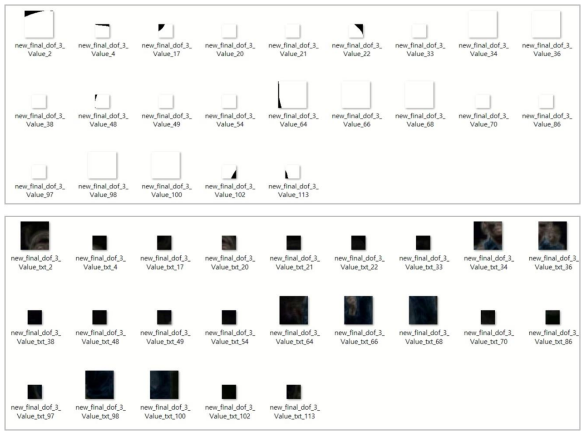


Fig. 5. Collected DoF patches in leaf nodes of quadtree (DoF patches : below images).

3. Convolutional Neural Network with DoF Patches

우리는 앞에서 설명한 D_p를 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 $\frac{1}{2} \| D - f(x) \|^2$를 최소화하는 것이다. 이 방정식을 최적화시키기 위해 우리는 아래와 같은 SRCNN[3]기반의 합성곱 신경망 방식을 사용하였다 (Fig. 6 참조).

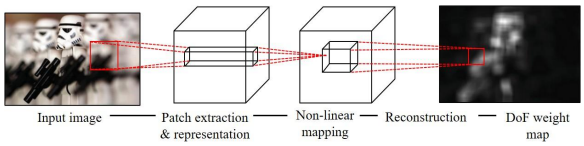
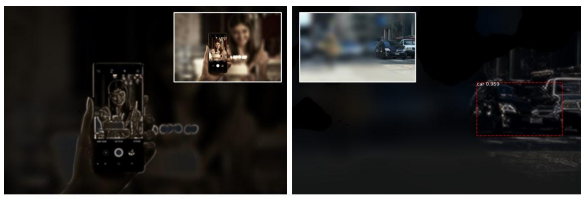


Fig. 6. DoF CNN(convolutional neural network).

Fig. 3에서 보여준 네트워크 방식을 적용하기 위해 입력 이미지의 DoF 가중치 맵에서 DoF 가중치 맵을 스무딩한 이미지로 변환하기 위한 가중치 매개변수를 계산해야 한다 : δ : → D : -(δ^* : → D^* : , θ). 여기서 δ, D는 입력 이미지와 그에 해당하는 DoF가중치 맵이며, δ^*, D^*는 δ를 스무딩한 입력 이미지이며, 그에 해당하는 DoF 가중치 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 진행했으며 좀 더 자세한 설명은 이전 DoF 네트워크 기법을 살펴보길 권장한다[2].

III. Conclusions

본 연구에서는 퀴드트리를 활용하여 DoF영역을 찾는 알고리즘을 최적화했고, 이 기술을 인공지능을 통해 빠르고 효율적으로 학습하고 테스트할 수 있는 새로운 합성곱 신경망 기법을 제안했다.



(a) Training image (before : 99 hours, our method : 18 hours) (b) Test : object detection

Fig. 7. Training and test results with our method (inset image : input image).

Fig. 7a는 본 논문에서 제안한 방법을 통해 얻어진 학습 결과이다. 학습 과정에서 425개의 적은 데이터들을 이용했다. 초기에는 검정색인 이미지에서 점차 수렴하면서 원본 이미지를 찾아가는 것을 볼 수 있으며, 특히 DoF로 인해 포커싱된 영역 주변으로 수렴해가는 결과를 얻었다. 학습 과정에서 퀴드트리를 이용하지 않았을 때는 99시간이 소요되었지만, 제안하는 방법은 18시간 만에 충분한 학습결과를 얻었다.

Fig. 7b는 학습된 결과를 객체 인식에 테스트해본 결과이며, 이미지에 포함되어 있는 모든 객체에 대해서 인식이 되는게 아니라 포커싱된 자동차 객체에 대해서 객체 인식이 이뤄진 것을 볼 수 있다.

Fig. 8은 다양한 이미지들에서 안정적으로 DoF 가중치 맵이 추출된 결과이다. 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.

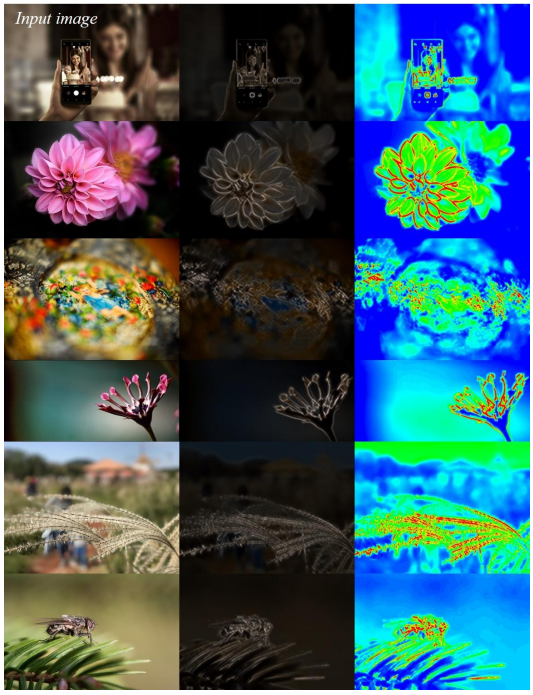


Fig. 8. Test results with our method. Right images are colored according to the magnitude of DoF weight map from red(high) to blue(low) (left : input image, middle : DoF weight map).

REFERENCES

[1] Gatys, Leon A., Alexander S. Ecker, and Matthias Bethge, Image style transfer using convolutional neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2414-2423. 2016.

[2] Donghui Kim, Jong-Hyun Kim, Convolutional neural network technique for efficiently extracting depth of field from images. Proceedings of the Korea Society of Computer and Information Conference, pp. 429-432, 2020.

[3] 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.