다목적 애플리케이션을 위한 피사계 심도 기반 후처리 프레임워크

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

Depth-of-Field based Post-Processing Framework for Multipurpose Applications

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

● 요 약 ●

본 논문에서는 합성곱 신경망을 통해 학습된 DoF(피사계 심도, Depth of field) 네트워크 아키텍처를 이용하여 객체 인식, 시점 추적, 문자 인식, 비사실적 렌더링 등 다양한 애플리케이션에 적용할 수 있는 사후 필터링 기법에 대해 살펴본다. 일반적으로 영상은 포커싱과 아웃포커싱에 의해 사용자의 관심표현이 결정되며, 이를 이용하여 영상 내 중요도를 판단한다. 영상 내에는 수많은 콘텐츠들이 혼재되어 있기 때문에 사용자가 집중적으로 보고 있는 콘텐츠를 찾아내기 어렵다. 본 논문에서는 사용자가 흥미롭고 집중적으로 보고 있는 영역을 DoF 네트워크로 학습시키고, 이를 통해 이전 기법으로는 표현할 수 없었던 DoF 기반 객체 인식, 시점 추적, 문자 인식, 비사실적 렌더링을 효율적으로 표현해낸다.

키워드: 객체 인식(Object recognition), 시점 추적(Viewport tracking), 문자 인식(Character recognition), 비사실적 렌더링(Non-photorealistic rendering), 피사계 심도(Depth of field)

I. Introduction

영상 내 콘텐츠의 분석, 객체 검출 등 사용자의 ROI를 기반으로 데이터를 분석하는 과정에서 중요한 특징 중 하나가 이미지의 피사계 심도이다[1]. 사용자가 영상을 봤을 때, 영상에는 포커싱과 아웃포커싱에 의해 선명하거나 뿌옇게 표현되는 차이가 있으며, 이러한 차이는 공간상에서 색상을 미분했을 때 더 큰 차이를 보이게 된다.

우리는 이러한 특징을 이용하여 합성곱 신경망 기반 네트워크를 구성하고, DoF 영역을 효율적으로 찾을 수 있도록 학습하고, 다양한 애플리케이션(객체 인식, 시점 추적, 문자 인식, 비사실적 렌더링)에 적용할 수 있는 후처리 프레임워크를 제안한다. 이러한 기법을 계산하기 위한 본 연구의 기여도는 아래와 같다 :

- 인공신경망을 통해 다목적 애플리케이션에 활용 할 수 있는 DoF 영역 추출 기술
- DoF 영역의 정확한 인식을 위한 사후 필터링 기술
- 필터링 된 DoF 영역을 이용하여 다양한 애플리케이션에 적용할 수 있는 방법

II. The Proposed Scheme

1. Detection of DoF region in a single image

이 과정은 이전 기법에서 제안한 방법을 활용했으며[3], 본 논문에서 간단하게 리뷰를 하도록 한다. 우리는 이미지로부터 DoF 영역을 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 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).

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

이미지에 존재하는 DoF 영역을 추출하기 위해 본 논문에서는 몇 가지 가정을 한다 :

1) DoF 영역에서는 DoF를 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(Smoothing) 이미지 I^{rgb}_1의 RGB색상 채널의 차이를 기반으로 DoF 가중치 맵 D를 다음과 같이 계산한다 (수식 3 참조).

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

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

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. 수식에서 보듯이 DoF 영역의 가중치는 RGB채널 값을 3차원 벡터로 표현하며, 그 벡터의 크기를 통해 계산한다. Fig. 1은 입력 이미지를 통해 얻은 DoF 가중치 맵인 D이다. DoF에 의해 Fig. 1a에서 포커싱된 뿌연 영역의 가중치가 잘 표현되었다.

(a) Input image (b) $\boldsymbol{D}$

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

2. Convolutional Neural Networks with DoF regions

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[2]기반의 네트워크 방식을 사용하였다 (Fig. 2 참조).

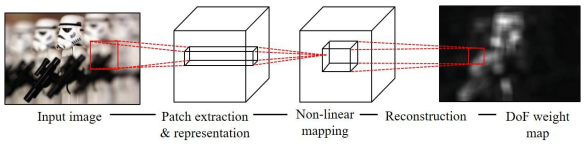
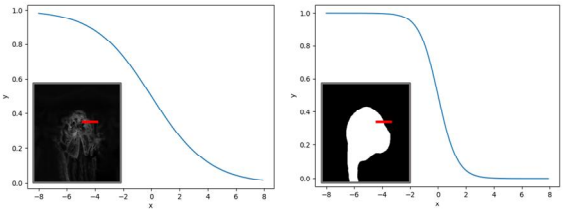


Fig. 2. DoF CNN(Convolutional neural network).

Fig. 2에서 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 DoF 가중치 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[3].

3. Post-Filtering to Calculate Binarized DoF regions

앞에서 설명한 네트워크를 이용하여 테스트를 하면 DoF 가중치 맵인 D를 계산할 수 있지만, 이 결과를 다양한 애플리케이션에 적용하기에는 어려움이 있다. 이 같은 문제가 풀기 어려운 이유는 두 가지이다. 첫 번째는 DoF 가중치에 대한 변화가 명확하지 않은 것이며 (Fig. 3a 참조), 두 번째는 DoF 영역에 대한 가중치가 부분적으로 다르기 때문에 객체 인식과 같은 애플리케이션에 정확도가 떨어지게 된다.



(a) Variation of $\boldsymbol{D}$ (b) Variation of $\boldsymbol{D^*}$

Fig. 3. Comparison result with variation of D and D^* in cross section (red).

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

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

여기서 f_{gm}과 x는 출력과 입력 이미지이며, M은 최댓값을 의미하고, 본 논문에서는 255를 사용하였다. 만약 g가 1이면 지수는 1이 되므로 선형적 밝기변화를 의미하며, 본 논문에서는 0.85를 사용하였다. 이렇게 얻어지는 f_{gm}의 색상은 좀 더 선명하지만, 경계나 에지부분에서 블러되는 문제가 있으며, 이 문제를 완화하기 위해 라플라시안(Laplacian) 커널을 이용한 영상 샤프닝(Sharpening)를 적용한다(수식 6 참조).

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

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. 3b 참조). Fig. 3a와 비교했을 때, 안정적으로 DoF 영역을 이진화 했으며, 단면에서도 영역 내/외부가 명확하게 차이가 나는 것을 볼 수 있다 (Fig. 3b 참조).

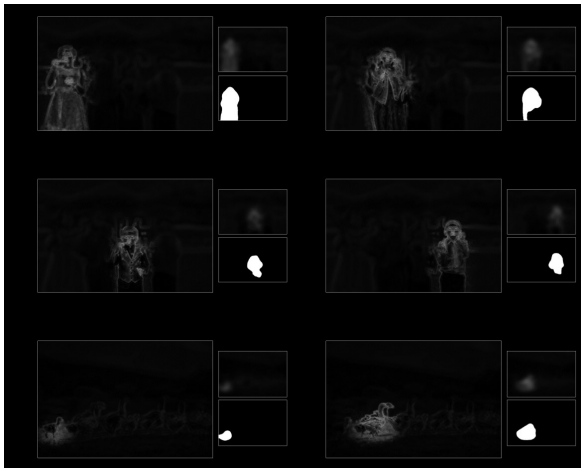
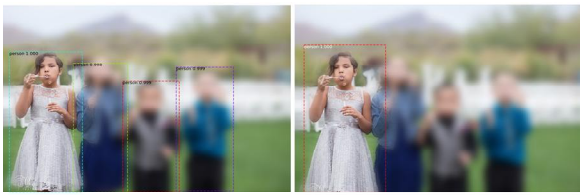


Fig. 4. Refined DoF weight maps from various input images.

Fig. 4는 D를 사후 필터링으로 거친 결과이며, 위에 있는 작은 이미지는 라플라시안 과정까지 적용한 결과이며, 아래 이미지는 영상 이진화 과정까지 적용한 결과이다. 그림에서 보듯이, D에서 가중치가 큰 영역인 흰색 영역을 부근으로 영상 이진화 처리를 안정적으로 수행하였다.

4. Validation Test

본 논문에서는 제안하는 방법을 통해 만들어진 D^{*}를 이용하여 유효성 검증을 위한 실험으로 4가지 애플리케이션에 적용을 해봤다.



(a) Without post-filtering (b) With post-filtering (our method)

Fig. 5. Object recognition.

첫 번째 실험은 D^{*}를 활용한 객체 인식이며, 우리는 YOLO(You only look once) 라이브러리를 이용했다. Fig. 5a에서 보듯이, 영상 내 DoF가 표현되었음에도 4명의 아이들이 모두 검출됐으며, 모두 "Person"이라고 인지했다. 객체 인지 정확도는 0.998 이상으로 상당히 높은 값을 나타냈다. 그에 반해, 우리의 방법은 포커싱된 아이만을 정확하게 인지했다 (Fig. 4b 참조).

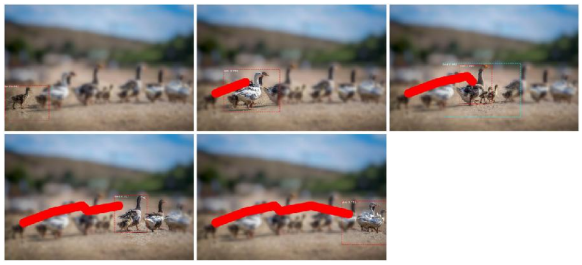


Fig. 6. Viewport tracking.

두 번째 실험은 DoF의 변화를 이용하여 시점을 추적한 결과이다 (Fig. 6 참조). 영상 내 DoF는 사용자가 집중적으로 보고 있는 특징을 표현하기 때문이 이를 이용하여 시점 변화를 정확하게 추적할 수 있다.

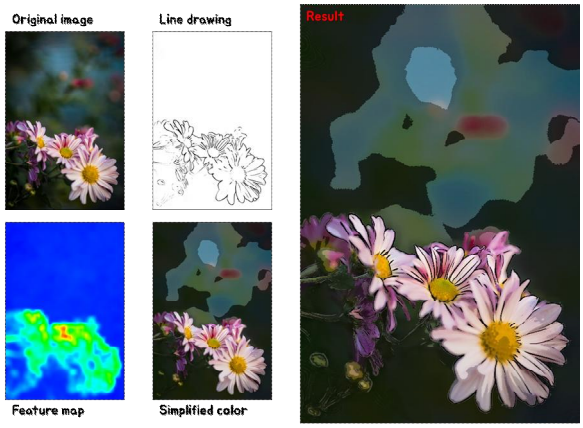
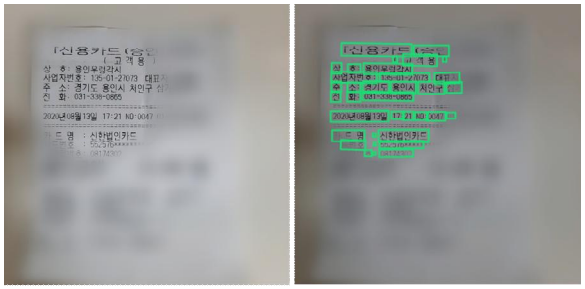


Fig. 7. Non-photorealistic rendering.

세 번째 실험은 D^{*}를 활용한 비사실적 렌더링이며 (Fig. 7 참조), 그림에서 보듯이 포커싱된 부분은 비사실실적 페인팅 레벨이 다소 낮지만, 아웃포커싱된 부분은 좀 더 강하게 색상 단순화가 적용되었다. 일반적인 비사실적 렌더링은 영상의 색상과 특징을 일괄적으로 단순화 시키기 때문에 원본 형태를 알아보기 어려울 때가 종종 있지만,

제안하는 방법은 D^*에 따라 색상 단순화 레벨을 자동으로 조정할 수 있기 때문에 이전기법들보다 훨씬 자연스러운 결과를 얻어낼 수 있다.

마지막 실험은 문자 인식인 OCR(Optical character recognition) 실험이다 (Fig. 8 참조). 앞에서의 실험과 마찬가지로 포커싱된 영역에 한해서 문자를 인식하기 때문에 사용자가 집중하고 있는 내용을 쉽게 인지할 수 있다.



(a) Input image (b) OCR test

Fig. 8. Character recognition.

III. Conclusions

본 연구에서는 DoF 영역의 사후 필터링을 활용하여 다양한 애플리케이션에 적용 할 수 있는 알고리즘을 제안했다. 기존의 인공신경망 기법으로도 DoF 영역을 안정적으로 추출할 수 있지만[3], 객체 검출, 인식 등 좀 더 명확한 분석을 위해서는 DoF 영역이 이진화로 되어 있어야 하며, 본 논문에서는 이를 안정적으로 표현할 수 있는 후처리 프레임워크를 제안했다. 향후, 우리는 DoF 가중치 맵을 활용하여 사용자 시점을 분석하고, 이 시점 변화로부터 나타나는 콘텐츠 해석이나 관심 영역 분석 등에 대해 추가 연구할 계획이다.

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] 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.

[3] 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.