

양방향 특징 피라미드를 이용한 그림자 객체 탐지

신석경 *

Instance Shadow Detection Using Bidirectional Feature Pyramid

Shin SeokGyeong*

요 약

새로운 Feature Maps을 계산하여 Instance Detection 기법에 적용해 그림자를 검출한다.

Abstract

This report presents a survey and a comparative evaluation of recent techniques for detecting shadow detection. Then by combining existing techniques, we implemented deep learning methods for shadow detection.

Key words

deep learning, instance segmentation, convolutional neural network, shadow detection

Github URL

<https://github.com/jiku100/ShadowDetectionMethod.git>

1. 서 론

본 연구는 컴퓨터 비전 기술을 통해 단일 영상에서 그림자를 검출하는 방법에 대해 설명한다. 컴퓨터 비전의 많은 응용 분야에서는 Detecting(탐지)와 Tracking(추적)을 사용하고 있다. 그림자의 존재는 이 2가지의 기능에 방해가 되므로, 검출된 그림자를 이용해 이미지로부터 방해가 되는 그림자를 제거할 수 있다. 또한, 검출된 그림자의 형태 및 구조를 분석하여 객체에 대한 더 정확한 정보를 추출할 수 있다. 특히, 그림자는 광원에 의존적으로 발생하는 현상이므로 그림자를 분석하여 광원의 세기, 위치 정보를 알 수 있다. 이러한 정보는 Computer

Graphic 분야와 Augmented Reality 분야에서 그림자를 생성하는 과정에서 사용될 가능성이 있다.

현재까지 그림자를 검출하기 위한 여러 방법들이 연구되어 왔다. 이 연구들을 크게 2가지로 분류하면 먼저, 특징을 직접 계산해서 그것을 이용해 그림자를 검출하는 Feature-Base 방법([1], [2], [3], [4], [5], [6], [7], [8])이 있고, 딥러닝 네트워크를 사용한 방법([9], [10], [11], [12])이 있다. 2016년 이후부터 딥러닝 네트워크 방법이 기존의 방법보다 정확도가 높아졌기 때문에 본 연구도 딥러닝 네트워크를 기반으로 하여 그림자 검출의 문제를 해결하였다.

본 연구는 영상 전체에 존재하는 그림자를 검출

* 중앙대학교

하는 것보다 객체와 그림자 간의 관계성을 찾아내고, 얻어진 Segment 사이의 계산을 통해 광원의 방향을 찾는 것을 목표로 한다.

II. Shadow Features

이번 장에서는 먼저 기존의 Feature-Base 방법들에 대한 분석을 진행한다.

1. Intensity

그림자는 일반적으로 어둡다는 특징을 가지기 때문에 Pixel 간의 Intensity difference를 계산하는 것으로 찾을 수 있다. 이때 Intensity difference는 pixel끼리 계산하는 방법과 Segment끼리 비교하는 방법이 존재하는데, Pixel 끼리는 absolute difference를, Segment끼리는 histogram 사이의 L1 norm을 통해 계산한다.

2. Chromaticity

그림자는 주변의 배경과 색채 간의 유사성을 지닌다. Pixel의 CbCr 값을 계산하고, 두 Pixel 간의 각도를 계산하는 것으로 Chromaticity distance를 계산할 수 있다.

3. Entropy Minimization

그림자의 특징을 사용하는 방법 이외에 영상 전체적으로 Illumination-invariant하게 변형하는 방법도 존재한다. RGB 값을 log scale로 변형하면 2차원 평면에서 선형성을 가지는 것을 확인할 수 있다. 이 값들을 Entropy가 최소화되는 방향으로 사영하면 영상에 존재하는 빛의 영향을 최소화할 수 있다.

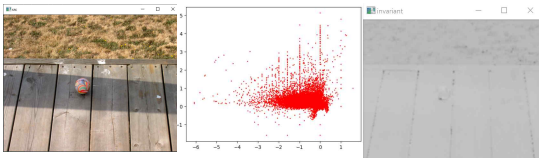


그림 1. Entropy Minimization 결과

4. Edge

그림자는 명암의 대비가 발생하는 공간이기 때문에 Edge가 명확하다. 이때 Edge의 특성을 계산하여서 그림자를 검출할 수 있다. 영상에 존재하는 Edge를 기반으로 특징을 계산하고, Decision tree를 사용해 그림자를 검출할 수 있다.[6]

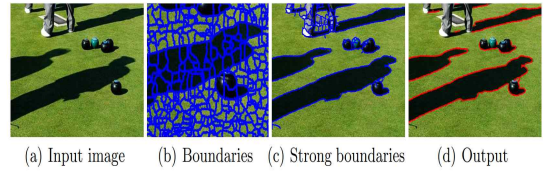


그림 2. Edge 기반 그림자 검출 결과[6]

5. Others

이외의 다른 특징들이 존재한다. 그림자는 배경에 비해 local variation을 줄이기 때문에 상대적으로 Smooth하다. 따라서 Smoothness를 계산하면 그림자 영역에서 상대적으로 작게 나온다.

비슷한 맥락으로 그림자에서는 비대칭도의 평균이 그림자가 아닌 부분보다 작게 검출된다.

이외에도 그림자의 경계에서 Gradient Value의 분포가 invariant한 Gradient Similarity, Texture properties가 배경과 유사하다는 Texture Similarity가 존재한다.

III. Feature-Base Method

앞서 언급한 특징들을 계산하여서 Shadow Detection에 활용할 수 있다. 일정 기준을 만족하는 Feature의 개수를 통해 해당 Pixel이 Shadow인지 판단하여 Shadow를 검출한다(그림 4). 입력된 이미지를 전처리 후 하나의 픽셀마다 특징 검사를 수행한다. 해당 Pipeline에서 사용한 특징은 CbCr distance, Y distance(Intensity), Smoothness, Local Max, Gradient Sum이다. 총 5개의 특징을 계산한 다음 3개 이상의 조건이 만족한다면 해당 픽셀을 그림자로 판단한다. 초기에 사용자 입력(Mouse Click)을 통해 기준 픽셀을 설정하고 해당 기준을 바탕으로 조건들을 설정한다. 모든 조건은 각각의 histogram을 구성한 이후, 1번째로 나오는 valley의 값을 threshold로 설정하여 threshold 이하일 경우 해당 조건을 만족한다고 판단하였다.

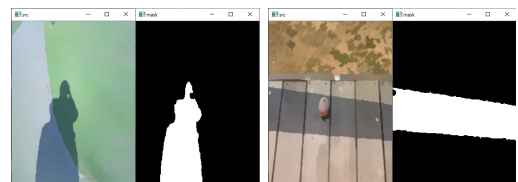


그림 3. 그림자 검출 결과

그림 3의 결과를 살펴보면, 그림자가 2개 이상의 surface에 존재하는 경우에도 올바른 그림자를 검출하고, 그림자 내부에 색깔이 다른 객체가 존재하여도 그림자를 검출하는 것을 확인할 수 있었다.

그림 4의 Pipeline은 기존의 그림자 검출 방법을 혼합하여 사용했기 때문에 존재하고 있던 여러 문제점을 해결한 결과를 얻을 수 있었다.

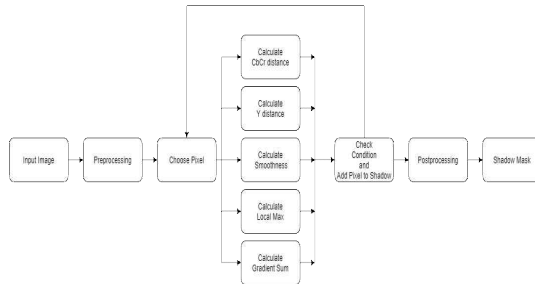


그림 4. Shadow Detection Pipeline

그러나 Feature-Base Method는 사전에 그림자의 특징이 무엇인지 개발자가 직접 알아내야 한다는 점, 조건들이 영상의 성질과 환경마다 다르기 때문에 다양한 상황에서 적용이 불가능하다는 단점이 존재한다. 또 계산과정에서 더 정확한 계산을 위해 연산량이 증가하게 된다면 상대적으로 속도가 느리다는 단점이 존재한다.

IV. Deep Learning Method

2016년도부터 앞서 설명한 Feature-Base Method를 넘어서 Deep Learning Network, 특히 CNN(Convolutional Neural Network)를 사용한 Shadow Detection Method가 많이 연구되고 있다. 가장 처음으로 연구된 방법은 영상의 정보를 가지고 Convolution Network를 훈련시키는 방법이 존재한다.[9] 이 방법은 영상으로부터 SuperPixel과 Edge를 구한 다음, 해당 부분의 Patch로 CNN을 학습시켜 Shadow Region을 구하는 방법이다. 이 방법은 기존의 Feature-Base의 방법과 Deep Learning 방법을 혼합한 방법으로 좋은 성능 향상을 이끌었다. 그러나 이 방법은 개발자가 먼저 계산을 해줘야 하고, 검정색 물체와의 구별이 되지 않는다는 단점이 존재하였다.

이후에는 입력 이미지만을 가지고 Network를 사용하는 End-to-End 방식의 CNN Network가 개발되었다. 그 중 BDRAR Network[10]는 Backbone에서 계산된 Feature Maps을 혼합하여 새로운 Feature Map을 계산하는 방법을 사용한다. 이 과정은 High Resolution에서 Low Resolution 방향으로 Feature Map을 혼합하는 방향과 Low Resolution부터 High Resolution으로 혼합하는 방향을 가지고 있어 양방향성을 가진다고 말한다.

두 개의 Feature Map을 혼합할 때에 Attention Map이라는 개념을 도입하는 것으로 Weighted Sum을 구현하였다. 해당 연구는 기존의 문제였던 검정색 물체에 대한 구별과 그림자 검출의 정확도에 대한 성능을 향상시켰다.

이후에는 CNN Network의 구조는 많이 변하지 않았지만 학습 방법을 새롭게 도입하여 훈련시킨 연구가 발표되었다. MTMT Network[11]은 Multi-Resolution Feature Map을 통해 Shadow를 검출하는 방법은 동일하지만, 학습 과정에서 준지도 학습 중 Mean Teacher 방식을 적용하여 더 많은 데이터를 훈련할 수 있었다. 또한 Shadow Region만 구별하는 것이 아니라, Shadow Edge, Shadow Count와 같이 그림자와 관련된 다른 정보도 추가로 학습시키면서 그 정확도를 향상시켰다.

본 연구자는 앞서 설명한 2가지의 Network를 혼합하였고, 입력받은 RGB 영상을 그대로 사용하는 것이 아니라 RGB 3 Channels과 Log Chromaticity 2 Channels, GrayScale 1 Channel, 총 6 Channels을 입력으로 하여 훈련을 진행하였다.

그림 5는 Channel의 개수나, Output의 종류를 달리하며 훈련을 진행한 결과이다. 3행의 결과를 보면, 6 Channels에 Edge, Count를 모두 훈련한 결과(EDGE + Count 열)가 다른 방법에 비해 어두운 아스팔트 부분과 그림자를 더 잘 구별한다는 것을 확인할 수 있다. 이는 2행의 결과도 비슷하게 보여준다. 그러나 1, 4행의 결과를 보면 오히려 배경과 그림자를 구별하지 못하는 것을 확인하였다.

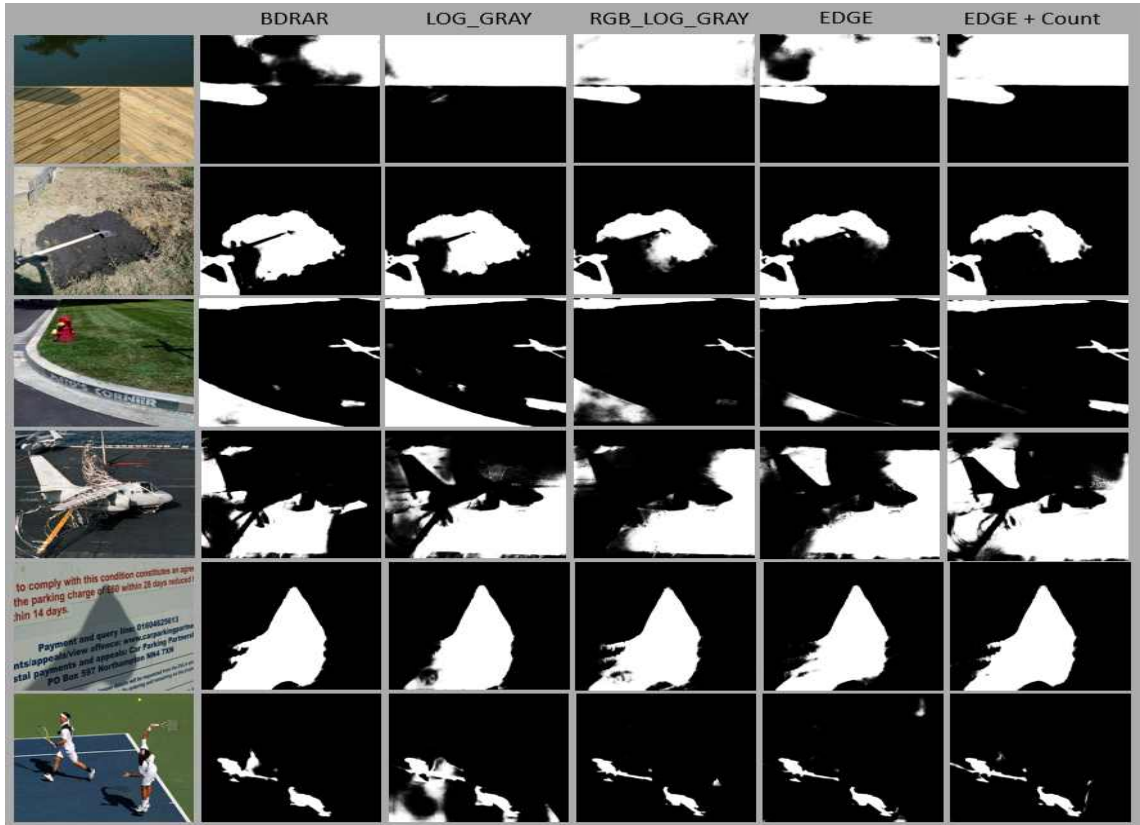


그림 5. BDRAR + MTMT Result

V. Instance Shadow Detection

지금까지의 방법은 영상 전체에 대해 그림자를 검출하는 방법이다. 그러나 2020년에 Mask R-CNN 이라는 새로운 방법이 제시되었다. 이는 Instance Segmentation 기법을 사용해서 객체와 그림자를 찾고, 그 사이의 연관성까지 찾는 Network이다.[12] 해당 연구에서는 Shadow 자체를 어떤 하나의 물체라고 판단하여 Detection에 활용하였다.

그러나 해당 연구는 Shadow Detection 자체에 목적을 두지 않았고, 객체 탐지 Network를 통해 Shadow를 찾았기 때문에 기존에 발표되었던 연구보다 상대적으로 Shadow Detection의 성능이 좋지 못하였다. 따라서 본 연구자는 해당 논문에서 소개된 LISA(Light-guided Instance Shadow-object Association) Network를 추가적인 수정을 통해 Shadow Detection의 성능을 향상시켰다.

VI. BDLISA

본 연구자가 선택한 방법은 위에서 설명한 양방향으로 Feature Map을 계산하는 방법을 LISA Network의 Backbone에 적용하는 방법이다. LISA Network의 Backbone은 Resnet-101의 Feature Pyramid Network 버전을 사용한다. 총 5개의 Feature Map이 출력되는데 이 5개의 Feature Map을 가지고 새로운 Feature Map을 계산하였다.

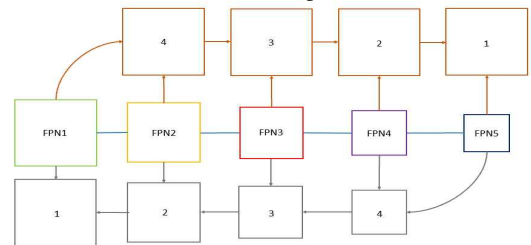


그림 6. BDLISA Backbone

그림 6을 통해 새로운 Backbone의 구조를 설명

한다. FPN으로부터 총 5개의 Feature Map이 계산되면, 해당 5개의 Feature Map을 가지고 양방향으로 새로운 Feature Map 8개를 계산한다. 이후 1번 Feature Map끼리 혼합, 2번 Feature Map끼리 혼합하여 총 4개의 새로운 Feature Map을 계산한다. 이때 혼합은 Attention Map을 사용하여 Weighted Sum을 통해 계산한다. 이후 FPN1의 Feature Map + 4개의 새로운 Feature Map을 가지고 나머지 Mask R-CNN 기법을 적용하여 Instance Segmentation을 진행한다. 즉, 양방향성을 고려한 새로운 Backbone으로 Mask R-CNN을 학습하는 것이다.

따라서 이 Network의 이름은 Bi-Directional Light-guided Instance Shadow-object Association, BDLISA로 결정하였다.

VII. Training

훈련 과정은 Facebook에서 개발한 최신 Mask R-CNN 기법인 detectron2의 과정을 따랐다. Ubuntu 18.04에서 Tesla P-100-PCIE-16GB의 그래픽 카드를 기반으로 훈련을 진행하였다. 총 48시간 동안 140,000 Epoch을 훈련하였다. 데이터셋의 경우 LISA에서 사용한 SOBA 데이터셋을 활용하였다. 해당 데이터셋은 840개의 기본 이미지가 존재하고 해당하는 Object Instances가 2,999개 Shadow Instances가 2,999개 존재한다.

이미지 개수	객체 개수	그림자 개수
840	2,999	2,999

표 1. 데이터 세트

특히 detectron2이 CUDA 10버전을 사용하기 때문에 CUDA 10 버전이 필수적이다.

VIII. Result

훈련을 통해 얻어진 가중치를 가지고 기존의 LISA Network와 비교하였다. 기존의 LISA Network의 문제는 그림자를 세밀하게 검출하지 못한다는 것이다. 즉, 그림 7에서 LISA의 결과를 보면 실제 그림자 영역보다 더 크게 그림자 영역을 검출하는 것을 확인할 수 있다. BDLISA Network는 이러한 문제를 해결하여 보다 더 정교하게 그림자를 검출할 수 있었다.



그림 7. Result (Left: BDLISA, Right LISA)

그림 7의 결과를 보면 전반적으로 그림자를 넓게 검출하는 LISA보다 BDLISA의 그림자 검출 면적이 더 정확한 것을 확인할 수 있다. 특히 2행의 결과를 보면 버스 그림자의 가장 오른쪽 부분이 LISA의 경우 더 넓게 검출하지만 BDLISA의 경우 그림자의 끝에 맞춰서 검출하는 것을 확인할 수 있다. 또한, 1행의 결과를 보면 자동차 뒷바퀴에도 그림자가 존재하는데 LISA의 경우 검출하지 못했지만, BDLISA는 검출하였다. 마지막으로 5행의 결과를 보면 LISA의 경우 그림자를 제대로 검출하지 못했지만, BDLISA의 경우 그림자를 정확하게 검출하는 결과를 보였다.

IX. Conclusion

본 보고서에서는 현재까지 연구된 그림자 검출 방법에 대해 살펴보고, 새로운 Network를 소개하였다. 기존의 그림자 검출 방식에는 Feature-Base Method와 Deep Learning Method가 존재하는데, Feature-Base의 경우 사전 지식의 필요성, 조건의

다양성, 연산의 복잡성과 같은 문제점이 존재하여 Deep Learning Method의 Shadow Detection이 최근 연구되고 있다. 특히, 본 보고서에서는 Instance Segmentation 기반인 LISA Network를 수정하여 좀 더 성능이 좋은 BDLISA Network를 소개하였다. 해당 연구는 일반적인 FPN의 Backbone에서 Bi-Directional Backbone을 사용하는 것으로 BDLISA Network를 개발하였다. 그 결과 기존의 LISA Network보다 좀 더 좋은 검출 성능을 확인할 수 있었다. 현재는 단순히 Backbone을 수정하는 것으로 그림자 검출 성능을 향상시켰지만, 이후에 Network의 구조를 개선하거나, Loss function을 개선하는 등 다양한 방법을 적용하는 것으로 Shadow Detection의 성능을 향상시킬 수 있을 것으로 보인다.

참 고 문 헌

- [1] Jiejie Zhu, Kegan G. G. Samuel, Syed Z. Masood, Marshall F. Tappen, Learning to Recognize Shadows in Monochromatic Natural Images, CVPR, 2010.
- [2] Ashraful Huq Suny1, Nasrin Hakim Mithila, A Shadow Detection and Removal from a Single Image Using LAB Color Space, IJCSI, 2013.
- [3] Yael Shor, Dani Lischinski, The Shadow Meets the Mask: Pyramid-Based Shadow Removal, EUROGRAPHICS, 2008.
- [4] Graham D. Finlayson, Mark S. Drew, Cheng Lu, Intrinsic Images by Entropy Minimization, ECCV, 2004.
- [5] Jung Yong Ju, Jan Young Woon, Choi Yun Woong, Cho Gi Sung, A Study for Introducing a Method of Detecting and Recovering the Shadow Edge from Aerial Photos, 한국측량학회지, vol. 24, no. 4, pp. 327-334, 2006.
- [6] Jean-Fran,cois Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan, Detecting ground shadows in outdoor consumer photographs, ECCV, 2010.
- [7] Andres Sanin, Conrad Sanderson, Brian C. Lovell, Shadow Detection: A Survey and Comparative Evaluation of Recent Methods, Pattern Recognition, 2012.
- [8] Jean-Fran,cois Lalonde, Alexei A. Efros, and Srinivasa G. Narasimhan, Estimating Natural Illumination from a Single Outdoor Image, ICCV, 2009.
- [9] S. H. Khan, M. Bennamoun, F. Sohel, R. Togneri, Automatic Feature Learning for Robust Shadow Detection, CVPR, 2014.
- [10] Lei Zhu1, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, Pheng-Ann Heng, Bidirectional Feature Pyramid Network with Recurrent Attention Residual Modules for Shadow Detection, ECCV, 2018.
- [11] Zhihao Chen, Lei Zhu, Liang Wan, Song Wang, Wei Feng, Pheng-Ann Heng, A Multi-task Mean Teacher for Semi-supervised Shadow Detection, CVPR, 2020.
- [12] Tianyu Wang, Xiaowei Hu, Qiong Wang, Pheng-Ann Heng, Chi-Wing Fu, Instance Shadow Detection, CVPR, 2020.