

반사 하이라이트 맵을 이용한 뉴럴 재조명

이연경[°] 고현성 이진우 김준호^{*}

국민대학교 비주얼 컴퓨팅 연구실

{dusrud1705, rh1363, jinwoolee, junho}@kookmin.ac.kr

Neural Relighting using Specular Highlight Map

Yeonkyeong Lee[°] Hyunsung Go Jinwoo Lee Junho Kim^{*}

Visual Computing Lab., Kookmin University

요약

본 논문에서는 사용자가 지정한 반사 하이라이트 맵을 가이드 영상으로 하는 뉴럴 재조명 기법을 제안한다. 제안하는 신경망은 다양한 조명 위치에서 렌더링된 영상을 이용해 사전학습시킨 백본 뉴럴 렌더러를 활용하며, 기저 영상과 렌더링 영상의 차이가 사용자가 제공한 반사 하이라이트 맵과 유사하도록 역전파에 의해 광원의 위치와 관련된 재조명 영상을 동시 최적화한다. 제안하는 방법은 아티스트가 선호하는 이차원 화면 공간 인터페이스를 제공하면서도 삼차원 조명의 위치를 명시적으로 추론할 수 있는 장점이 있다. 제안하는 뉴럴 재조명의 성능은 실제 값을 설정할 수 있는 실험 상황을 수립하여, 본 논문의 방법이 주어진 하이라이트 맵을 얼마나 잘 반영하는지 평가하고 실제 하이라이트 맵으로 추론한 조명 및 재조명 영상의 오차를 측정하였다. 제안하는 뉴럴 재조명이 추정된 광원 위치의 평균 오차율은 정규화된 삼차원 장면 크기 대비 0.11이다.

Abstract

In this paper, we propose a novel neural relighting that infers a relighted rendering image based on the user-guided specular highlight map. The proposed network utilizes a pre-trained neural renderer as a backbone network learned from the rendered image of a 3D scene with various lighting conditions. We jointly optimize a 3D light position and its associated relighted image by back-propagation, so that the difference between the base image and the relighted image is similar to the user-guided specular highlight map. The proposed method has the advantage of being able to explicitly infer the 3D lighting position, while providing the artists' preferred 2D screen-space interface. The performance of the proposed network was measured under the conditions that can establish ground truths, and the average error rate of light position estimations is 0.11, with the normalized 3D scene size.

키워드: 뉴럴 재조명, 뉴럴 렌더러, 반사 하이라이트 맵, 광원 위치 추정

Keywords: Neural Relighting, Neural Renderer, Specular Highlight Map, Light Position Estimation

1. 서론

컴퓨터 그래픽스의 발전으로 인해 실세계 장면과 거의 차이가 없는 가상 장면에 대한 사실적 렌더링(photo-realistic rendering) 영상을 얻는 것이 가능해졌다. 사실적 렌더링 영상을 얻기 위해서는 일반적으로 삼차원 장면을 구성하는 여러 가상 객체와 광원 간의 일어나는 빛의 상호작용을 전역 조명(global illumination) 알고리즘으로 근사하는 물리 기반 렌더링(physically-based rendering)을

활용한다 [1]. 이러한 물리기반 렌더링은 Maya, 3ds Max, Blender 등과 같은 삼차원 그래픽 저작 도구를 이용해 삼차원 장면을 구성하는 형상, 재질, 조명 조건 등을 각종 장면 요소를 사용자가 직접 조절하여 높은 품질의 영상을 만들어 낼 수 있는 장점이 있지만 방대한 양의 직간접 광원을 모두 고려해야 물리적으로 정확한 영상을 획득할 수 있기 때문에 오랜 렌더링 시간이 걸리는 단점이 있다.

비교적 짧은 렌더링 시간을 가지면서도 고품질의 영상을 얻

*corresponding author: Junho Kim/Kookmin University(junho@kookmin.ac.kr)

을 수 있는 사실적인 렌더링 방법의 일환으로 뉴럴 렌더링(neural rendering) 기법에 관한 연구가 최근 활발히 진행되고 있다 [2]. 뉴럴 렌더링은 다수의 이차원 렌더링 영상을 통해 삼차원 장면의 은닉 정보를 학습하는 뉴럴 렌더러(neural renderer)를 훈련시킨 다음, 훈련된 뉴럴 렌더러를 이용해 훈련에 등장하지 않았던 조건에 대응하는 이차원 렌더링 영상을 추론하는 방식으로 동작한다. 뉴럴 렌더러가 학습한 삼차원 장면의 렌더링 정보는 신경망의 텐서 구조로 은닉되어 표현되기 때문에, 삼차원 물체의 형상, 빛의 위치, 물질의 재질 정보 등이 명시적인 형태로 표현되지 않는 것이 특징이다.

본 논문에서는 사용자가 제공하는 반사 하이라이트 맵을 가이드 영상으로 하는 뉴럴 재조명(neural relighting) 기법을 제안한다. 삼차원 장면에 대한 초기 렌더링에 해당하는 기저 영상(base image)이 주어지면 사용자는 반사 하이라이트가 비치길 원하는 부분을 화면 공간에 맵 형태로 그린다. 제안하는 뉴럴 재조명 기법은 사용자가 제공한 반사 하이라이트 맵이 반영된 최종 렌더링 결과물과 함께 해당 하이라이트를 발생시키는 광원의 삼차원 위치를 동시 추정(joint estimation)한다.

이를 위해 다양한 조명 위치에서 렌더링 된 영상을 통해 학습된 백본 뉴럴 렌더러(backbone neural renderer)를 기반으로 기저 영상과 렌더링 결과물의 차이가 사용자가 가이드한 하이라이트 맵과 유사하도록 역전파(back propagation)를 수행함으로써 명시적인 조명 위치가 추론되도록 신경망을 설계하였다.

제안 네트워크의 성능을 평가하기 위해 두 가지 실험 상황을 수립하였다. 먼저 본 논문의 프레임워크가 사용자의 하이라이트 맵을 얼마나 잘 반영하는지를 측정하기 위해 테스트셋에 있는 렌더링 영상과 기저 영상의 차이를 실제(Ground Truth, GT) 값으로 간주하고 실제 조명 위치를 추정하는 기준 방법(baseline) 네트워크를 학습하였다. 이 기준 방법 네트워크가 사용자 하이라이트 맵에 대해 추정한 조명 위치로 렌더링한 결과와 기저 영상의 차이 맵과 사용자 하이라이트 맵의 PSNR을 측정하였다. 같은 방식으로 본 논문의 프레임워크가 추정한 재조명 영상과 기저 영상의 차이 맵과 사용자 하이라이트 맵의 PSNR을 측정하고 비교해 본 프레임워크가 사용자의 의도를 반영한다는 사실을 보였다. 또한 본 프레임워크의 정확도를 측정하기 위해 실제 하이라이트 맵으로 조명 위치와 재조명 영상을 추론한 뒤, 이 추론 결과들의 오차 값을 비교하였다. 실험 결과 추정 광원의 위치 오차율은 삼차원 장면의 크기 대비 0.11 오차율을 가진다.

제안하는 방식은 일러스트레이터 아티스트들이 이차원 화면 공간에서 음영 조절을 하는 방식과 매우 유사한 인터페이스를 가지면서도 기존 렌더링 소프트웨어에서 적용할 수 있는 삼차원 광원 위치를 직접 추론할 수 있다는 장점이 있다. 따라서 삼차원 렌더링 소프트웨어에 익숙하지 않은 일반 사용자뿐만 아니라 전문가들에게도 자신이 원하는 고품질의 렌더링 결과에 쓰일 광원의 위치와 그에 따른 렌더링 결과물을 예측하는데 활용될 수 있을 것으로 기대한다.

2. 관련 연구

2.1 뉴럴 렌더러

딥러닝의 발전으로 삼차원 장면에 대한 렌더링 영상을 생성 네트워크를 통해 만들어 내는 방법들이 새롭게 제안되고 있다 [2]. Sakurikar 등 [3]은 단일 시점으로 바라본 삼차원 장면의 물체에 대해, 재질이나 빛의 위치 변화에 따라 해당 물체에 대한 광선 추적(ray tracing) 렌더링 영상의 근사값을 얻어낼 수 있는 뉴럴 렌더러를 제안하였다. 이 렌더러는 시간이 오래 걸리는 광선 추적 렌더링을 네트워크를 통해 추론하여 빠른 시각화를 가능하게 한다. 이를 위해 재질과 조명의 시각화와 편집이 가능한 대화형 도구를 개발하였고, 광선 추적 렌더링 영상을 포함한 대규모 데이터셋을 공개하였다.

Thies 등 [4]은 뉴럴 텍스처(neural texture)를 이용한 지연 렌더링(deferred rendering)으로 새로운 시점에 대한 영상을 합성할 수 있는 뉴럴 렌더링 기법을 제안하였다. 이 방법은 삼차원 장면에 있는 물체의 텍스처 공간으로부터 최종 화면에 이르는 지연 렌더링 파이프라인을 신경망을 도입하여 구축한 다음, 주어진 삼차원 장면을 여러 각도로 바라본 훈련 영상으로부터 역전파를 통해 물체의 뉴럴 텍스처를 학습하는 방식으로 이루어져 있다. 이렇게 학습된 뉴럴 텍스처는 물체의 디테일한 형상, 재질, 빛 조건 등에 대한 정보를 함축적으로 은닉하고 있기 때문에 물체의 프록시 삼차원 형상(proxy 3D geometry)만으로도 새로운 시점에 대한 사실적인 합성 결과를 얻을 수 있다.

Zhang 등 [5]은 영상을 환경 조명(environment lighting)과 물체의 내재적인 특성(object intrinsic attribute), 빛 전달 함수(light transport function)의 세 가지 요소가 상호조합된 결과로 해석하고, 렌더링 영상 데이터를 기반으로 이들 요소를 학습하는 뉴럴 렌더러를 제안하였다. 제안하는 뉴럴 렌더러와 기존 물리 기반 렌더러를 통합해 재조명뿐만 아니라 새로운 시점에 대한 영상 합성도 가능함을 보였다.

본 연구는 삼차원 장면에 대한 다양한 렌더링 조건을 신경망의 텐서 형태로 학습한 뉴럴 렌더러를 백본 네트워크로 활용하여 사용자가 입력으로 준 반사 하이라이트 맵을 이용해 광원의 위치를 직접적으로 추정한다는 점에서 기존 연구들과 차별점을 가진다.

2.2 영상 기반 광원 추정

뉴럴넷을 통해 주어진 영상에 대한 물체의 표면 재질과 빛의 방향을 알아내는 연구가 활발히 진행되었다. 이러한 연구들은 라이트 프로브(light probe) 등과 같은 중간 단계의 광원 표현을 알아내는 방법 [6, 7, 8, 9]과 이러한 중간 단계 표현 없이 빛의 위치나 방향을 바로 알아내는 방법 [10]으로 나눌 수 있다.

Gardner 등 [7, 6]은 실세계의 구조화되지 않은 조명들을 파라미터화한 후 이미지로부터 정의한 파라미터를 추론하는 네트워크를 제안하였다. Wang 등 [8]은 같은 재질이 여러 조명에서 관측된 영상들과 다른 재질이 하나의 조명에서 관측된 영상들을 행렬

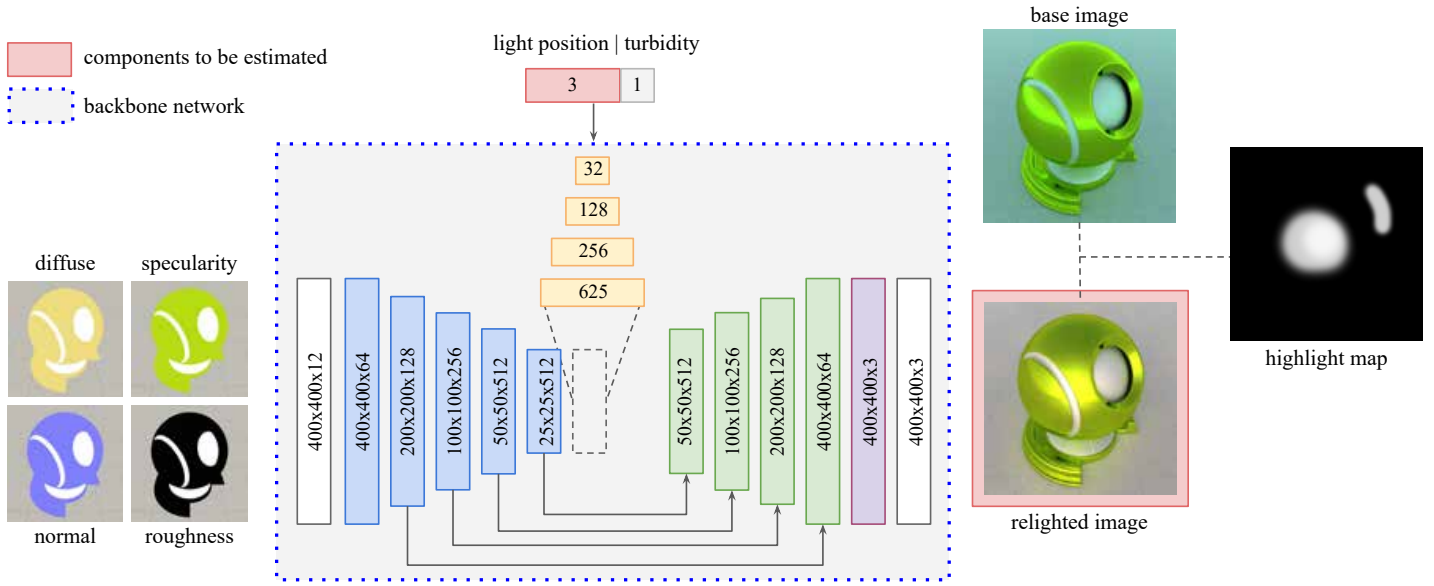


Figure 1: Architecture of the proposed framework.

구조의 정형화된 훈련 데이터로 모아 물체의 재질과 조명을 동시에 추정하는 신경망을 제안하였다. Zhou 등 [9]은 얼굴 사진들에 대해 알고리즘적인 방법으로 데이터셋을 만들고, 이를 이용해 재조명 처리된 인물 사진을 생성하는 네트워크를 제안하였다. Kán 등 [10]은 상대적 오일러 각도 표현(Euler angle representation)을 이용해 RGB-D 영상으로부터 3차원 좌표계 상의 조명 방향을 실시간으로 알아내는 방법을 제안하였다.

본 논문에서 제안하는 방법은 화면 공간에서 사용자가 하이라이트 맵을 가이드해 줄 수 있다는 점에서 기존의 영상 기반 광원 추정 방식에 비해 보다 효과적인 사용자 인터페이스를 제공한다. 이는 점에서 차별성을 가진다.

2.3 사용자 스트록에 기반한 재조명

사용자 스트록에 기반하여 주어진 영상의 셰이딩(shading)을 업데이트하는 연구가 스타일라이즈드 렌더링(stylized rendering) 분야에서 진행되었다 [11, 12, 13, 14]. Sloan 등 [11]과 Fišer 등 [14]은 사용자가 구체 등 간단한 기하도형에 적용한 셰이딩 예제로부터 해당 셰이딩이 복잡한 모델에 적용되었을 때의 결과를 만들어 내는 예제 기반(example-based shading) 기법을 제안하였다. 이 방법들은 단순한 기하 도형에 사용자가 셰이딩한 예제의 법선(normal) [11], RGB 색상 [13, 12], 빛 전달(light propagation) [14] 정보 등이 일치하는 패치를 복잡한 모델이 그려진 영상에서 찾아 셰이딩을 업데이트한다. 최근 Zhang 등 [15]은 아티스트가 부여한 스트록의 밀도로부터 광원 효과를 추정할 수 있다는 가정 하에, 스트록 밀도맵(stroke density map)을 생성하고 이를 이용해 이차원 영상을 재조명하는 방법을 제안하였다. 볼록 껍질 기반 색상 팔레트(convex-hull based color palette)를 사용하여 영상에서 사용된 색상 조합을 알아내고, 이를 기반으로 생성한 스트록 밀

도 맵을 일종의 높이 함수로 해석하여 주어진 광원의 위치에 대한 재조명 영상을 생성한다. 이 방법은 아티스트에게 친숙한 인터페이스를 제공하고 있지만, 데이터에 기반한 방법이 아니기 때문에 아티스트가 직접 셰이딩한 결과와는 다소 차이가 있다는 한계가 있다.

3. 프레임워크

본 논문에서는 기저 영상 위에 사용자가 지정(annotation)한 반사 하이라이트 맵을 이용해, 반사 하이라이트 맵을 반영하는 조명 위치와 재조명 영상을 추론한다. 이를 위해 [3]에서 제안하는 뉴럴 렌더러를 백본 네트워크로 활용하였다. 백본 네트워크의 모든 파라미터는 고정하고, Fig. 1에 빨간색 박스로 그려진 요소를 역전파로 변화하도록 만들어 최적화를 수행한다. 이 때, 뉴럴 렌더러의 은닉 정보는 렌더링된 영상과 기저 영상의 차이가 반사 하이라이트 맵과 유사하도록 최적화되며 사용자의 의도를 반영하는 명시적 조명 위치는 동시 최적화 된다.

3.1 백본 뉴럴 렌더러

본 논문에서 사용하고 있는 백본 뉴럴 렌더러는 재질의 시각화를 위해 고정된 형상 및 시점에 대해 다양한 재질과 조명을 적용한 광선 추적 기반 렌더링 결과를 학습하는 신경망이다 [3]. 고정된 형상에 대해 5,000개의 재질과 각각에 대한 5개의 랜덤 조명 위치를 하나의 시점으로 렌더한 영상들을 학습 데이터로 사용해 학습한 다음, 테스트 시에는 학습에 포함되지 않은 입력이 주어졌도 해당 입력에 대응하는 렌더링 결과를 빠르게 추론한다.

본 논문의 프레임워크에 입력으로 사용되는 렌더링 정보의 표현법은 뉴럴 렌더러 [3]에서 사용하는 것을 그대로 따른다. 재질

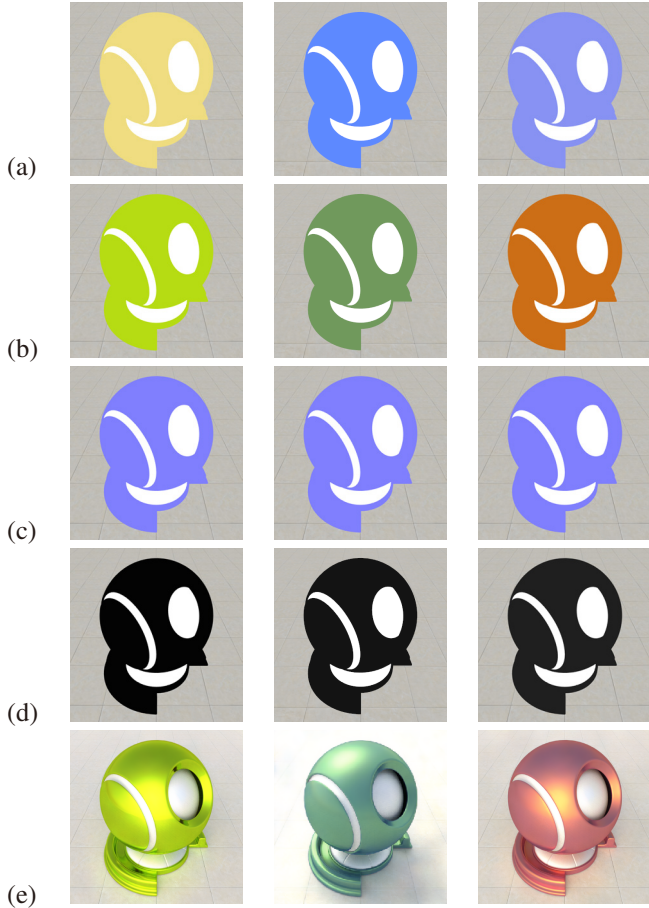


Figure 2: Input and output of neural renderer. (a) diffuse map, (b) specular map, (c) normal map, (d) roughness map, and (e) neural rendered image with light condition.

은 Fig. 2에서 보이는 것처럼 분산(diffuse), 반사(specular), 거칠기(roughness), 법선(normal)맵들의 조합으로 표현한다. 조명은 Hosek-Wilkie 하늘 모델(sky model) [16]을 사용해 환경 조명(environment light)을 조명의 3차원 위치와 흐림 정도(turbidity)의 조합으로 표현한다. 따라서 백본 네트워크는 해당 재질을 표현하는 4개의 맵과 조명을 표현하는 4차원 벡터를 입력으로 하여 뉴럴 렌더링된 영상을 출력한다.

본 논문에서는 이 뉴럴 렌더러 [3]를 제안하는 뉴럴 재조명 프레임워크의 백본 신경망으로 사용한다.

3.2 뉴럴 재조명

제안하는 뉴럴 재조명 프레임워크는 사용자가 지정한 반사 하이라이트 맵으로부터 이를 반영하는 Hosek-Wilkie 하늘 모델 [16]의 조명 위치를 최적화하고 대응되는 재조명 영상을 생성한다.

백본 뉴럴 렌더러의 입력으로 들어가는 조명 위치를 변경 가능하게 하여 백본의 은닉층으로부터 조명 위치를 명시적으로 추론할 수 있게 된다. 추론한 결과의 정확도는 재조명된 영상과 기저 영상의 차이와 유저의 하이라이트 맵에 대해 패치 기반 평균 제곱 오차(Mean Squared Error, MSE)로 측정한다.

네트워크 구성 Fig. 1에 보이듯이 우리의 프레임워크의 입력은 특정 재질에 대한 4 가지 재질 맵 (Fig. 2), 조명 정보, 기저 영상, 그리고 사용자가 화면 공간(screen-space) 인터페이스로 지정한 반사 하이라이트 맵이다. 이 영상들을 입력으로 받아 반사 하이라이트 맵을 반영하는 조명의 위치와 해당 조명을 이용해 재조명된 영상을 출력한다. 백본 신경망의 입력인 재질 맵과 조명 정보 중 조명의 위치를 변화할 수 있게 하였고 백본 신경망의 모든 파라미터는 고정하여 전체 프레임워크에서 조명의 위치만 역전파를 통해 변화하도록 하였다. 재질 맵과 조명 위치가 입력으로 주어지면 네트워크는 출력 이미지를 생성하고, 역전파를 통해 조명의 위치를 최적화한다.

사용자 가이드 반사 하이라이트 맵 Fig. 1에서 기저 영상(base image)은 특정 재질에 대해 조명 위치를 원점 $(0, 0, 0)^T$ 으로 설정하고 렌더링한 영상이다. 사용자가 기저 영상 위에 반사 하이라이트를 표시하면 사용자의 입력으로부터 Fig. 1의 하이라이트 맵(highlight map)을 획득한다. 우리의 프레임워크는 반사 하이라이트 맵을 만족하도록 재조명 영상과 조명의 위치를 최적화한다.

손실 함수 입력으로 모델의 재질에 대한 분산맵 \mathbf{D} , 반사맵 \mathbf{S} , 노말맵 \mathbf{N} , 거칠기맵 \mathbf{R} , 그리고 조명의 위치 \mathbf{l} 가 주어지면 뉴럴 렌더러 $\mathcal{F}(\cdot)$ 로 입력을 반영하여 이미지를 렌더링한다. 렌더링된 이미지에서 기저 이미지 \mathbf{B} 를 뺀 이미지가 반사 하이라이트 맵 \mathbf{A} 과 유사해지도록 조명의 위치 \mathbf{l} 가 최적화된다. 손실함수는 사용자의 지정한 반사 하이라이트 맵 \mathbf{A} 을 최대한 반영하기 위해 패치 기반 평균 제곱 오차(Mean Squared Error, MSE)를 사용한다. 이미지를 작은 패치 단위로 나누어 사용자의 지정한 반사 하이라이트가 지나가는 영역에 해당하는 평균 제곱 오차만 손실함수에 반영한다.

$$\mathcal{L} = \frac{1}{N_{pos}} \sum_{i=1} \mathbb{1}(\mathbf{A}_i) \|(\mathcal{F}_i(\mathbf{D}, \mathbf{S}, \mathbf{N}, \mathbf{R}, \mathbf{l}; \theta) - \mathbf{B}_i) - \mathbf{A}_i\|^2. \quad (1)$$

여기서 θ 는 뉴럴 렌더러 $\mathcal{F}(\cdot)$ 의 파라미터를 의미하고, 최적화 과정 전에 미리 학습되어 최적화 과정에서는 업데이트 되지 않는다. $\mathbb{1}(\cdot)$ 는 지시 함수로 반사 하이라이트 맵 \mathbf{A} 의 i 번째 패치 영역에 사용자가 지정한 반사 하이라이트가 있으면 1, 아니면 0을 출력한다. N_{pos} 는 사용자가 지정한 반사 하이라이트가 있는 패치의 개수이다.

4. 실험

4.1 구현과 학습

본 논문에서 제안하는 프레임워크는 뉴럴 렌더러¹ [3]를 활용해서 구성하였다. 네트워크의 입력 맵들과 출력 영상의 해상도는 400×400 픽셀이며, 조명 벡터의 차원은 3차원으로 하였다. 뉴럴

¹<https://github.com/AakashKT/NeuralMaterialVisualization>

Table 1: Quantitative evaluations with user annotation.

Method	shaderball	
	PSNR (dB) ↑	
	Mean	Med.
baseline	09.42	09.18
Ours	10.52	10.51

렌더러는 [3]에서 제안된 학습 방법으로 미리 학습 시켜 사용하였다.

주어진 반사 하이라이트 맵(specular highlight map)을 만족하도록 조명의 위치와 재조명 영상을 최적화하기 위해, Adam optimizer를 사용하여 0.001의 학습률(learning rate)을 사용하여 6,000번 반복하여 최적화한다. 미리 학습된 뉴럴 렌더러의 파라미터들은 변화되지 않도록 했으며, 조명의 위치만 최적화한다. 조명의 위치와 재조명 영상을 최적화하는 데 걸리는 시간은 NVidia GTX 2080을 사용한 환경에서 약 4분 정도 소요된다.

4.2 데이터셋

제안하는 프레임워크의 백본 뉴럴 렌더러를 학습하기 위해 [3]에서 제공한 셰이더볼(shaderball)과 더불어 유타 찻주전자(Utah teapot), 파라오(Pharaoh) 모델에 대한 장면 데이터셋을 각각 구성하였다. 백본 뉴럴 렌더러는 각 장면마다 별개로 학습되며, 각 장면에 대한 학습 데이터셋 구성은 다음과 같다. 데이터셋은 단일 시점과 형상에 대해 Fig. 2과 같이 물체 표면의 재질(material)을 구성하고, 여러 재질에 대해 랜덤한 위치에 설정한 광원으로 렌더링한 Fig. 2의 (e)와 같은 영상들을 제공한다. 데이터셋이 제공하는 재질의 수는 5,000개이며 각 재질에 대해 5개의 조명을 적용해 총 25,000개의 영상을 구성한다. 재질 영상과 렌더링된 영상의 크기는 400×400 이다. 추가로 각 재질에 대해서 광원 위치를 원점으로 설정하고 Blender [17]의 Cycle 렌더러로 렌더링하여 생성한 5,000개의 기저 영상을 본 연구의 성능을 평가하는데 사용하였다.

4.3 비교

제안하는 방법으로 뉴럴 재조명된 영상과 추정되는 조명의 위치가 사용자의 입력으로 만든 반사 하이라이트 맵을 얼마나 정확하게 반영하는지를 실험을 통해 확인한다.

기준 방법 제안하는 방법과의 비교를 위한 기준 방법(baseline method)으로 조명의 위치를 추정하는 네트워크를 구성하였다. 기준 방법은 기저 영상과 반사 하이라이트 맵을 입력으로 받아 광원의 위치를 추정한다. 데이터셋으로부터 기저 영상과 렌더링된 영상(GT image)의 차이 맵을 계산하여 Fig. 6(b)과 같은 실제 하이라이트 맵(GT highlight map)을 구성했다. 네트워크는 기저 영상과 실제 하이라이트 맵을 입력으로 받아 광원의 위치를 추정한다. 이를 위해 뉴럴 렌더러 [3]의 인코더 부분을 변경하여 구성

Table 2: Quantitative evaluations with GT annotation.

Method	light position		PSNR (dB)	
	Mean	Med.	Mean	Med.
Ours	0.11	0.13	28.89	28.46

하였다. 25,000개의 입력 영상과 조명 위치로 구성된 데이터셋을 이용하여, Adam optimizer를 사용해 30 epoch으로 학습시켰다.

평가 Fig. 3은 기저 영상(Fig. 3(a))을 사용자가 지정한 반사 하이라이트 맵(Fig. 3(b))을 이용해서 제안하는 방법으로 재조명된 영상(Fig. 3(d))과 추정된 조명의 위치를 이용해 렌더링한 영상(Fig. 3(e)), 그리고 기준 방법으로 추정한 조명의 위치를 이용해 렌더링한 결과(Fig. 3(c))를 보여준다. 제안하는 방법으로 재조명되는 영상과 추정되는 조명 위치의 결과가 기준 방법의 결과보다 하이라이트 맵을 잘 반영하는 것을 확인할 수 있다.

우리는 각 방법으로 추정한 영상과 기저 영상의 차이 맵과 사용자가 지정한 하이라이트 맵의 차이를 PSNR(Peak Signal-to-Noise Ratio)로 계산하였다. PSNR이 높을수록 추정한 결과가 반사 하이라이트 맵을 잘 반영한다고 할 수 있다. Table 1에서 제안하는 방법이 기준 방법과 비교하였을 때 높은 PSNR을 나타내는 것을 확인할 수 있다.

Fig. 4와 Fig. 5은 유타 찻주전자와 파라오 모델로 구성된 장면에 대해 동일한 실험을 수행한 결과이다. Fig. 4와 Fig. 5의 (a)와 (b)는 각각 해당 장면의 기저 영상과 사용자가 지정한 반사 하이라이트 맵을 보여주며, (c)와 (d)는 각각 제안하는 방법으로 재조명된 영상과 추정된 조명 위치의 결과를 이용해 Blender [17]의 Cycle 렌더러로 렌더링한 결과를 보여주고 있다. Fig. 4와 Fig. 5의 (c)와 (d)의 렌더링 결과가 서로 유사한 것을 확인할 수 있었으며, 이를 통해 본 논문에서 제안하는 뉴럴 재조명 기법이 사용자 지정 반사 하이라이트 맵을 반영하는 재조명 결과와 이에 해당하는 삼차원 조명 위치에 대한 동시 추정이 가능함을 재확인하였다.

4.4 조명 위치 추정

테스트 데이터셋 제안하는 방법이 추정하는 조명의 위치에 대한 정확도를 측정하였다. 먼저 실제 조명 위치에 대한 테스트 데이터셋을 구성하였다. Fig. 6(a)와 같이 기저 영상(base image)을 선택하고 기본 이미지를 렌더링한 재질과 임의의 위치에 대한 광원을 이용해 Fig. 6(c)와 같이 반사광이 적용된 영상을 렌더링하였다. Fig. 6(b)와 같은 반사 하이라이트 맵은 실제 조명의 위치를 적용해 렌더링한 영상(Fig. 6(c))과 기저 영상(Fig. 6(a))의 차이로 생성하였다. 실험을 위해 100개의 기저 영상, 실제 반사 하이라이트 맵, 그리고 실제 조명의 위치와 이로 렌더링한 실제 영상을 생성하였다. 테스트셋은 훈련 데이터셋과 마찬가지로 각 재질에 대해 Hosek-Wilkie 하늘 모델 [16]로 표현되는 조명 조건으로 Blender [17]의 Cycle 렌더러를 이용해 생성했다.

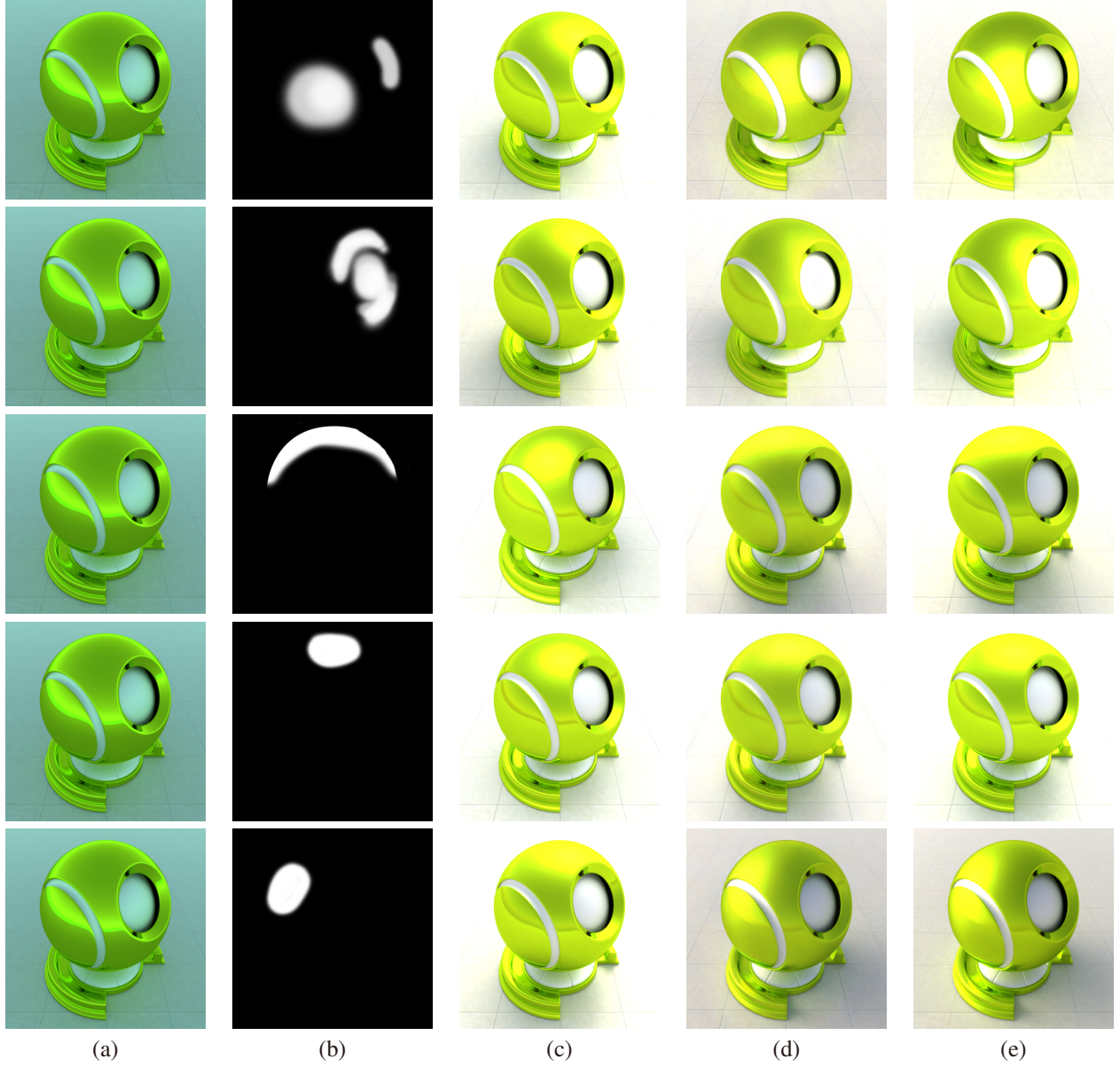


Figure 3: Result of our framework with the shaderball model. (a) base image, (b) user annotated specular highlight map, (c) rendered image with predicted light position using baseline method, (d) **our** neural relighted image, (e) rendered image obtained from a conventional 3D graphics renderer using **our** predicted light position.

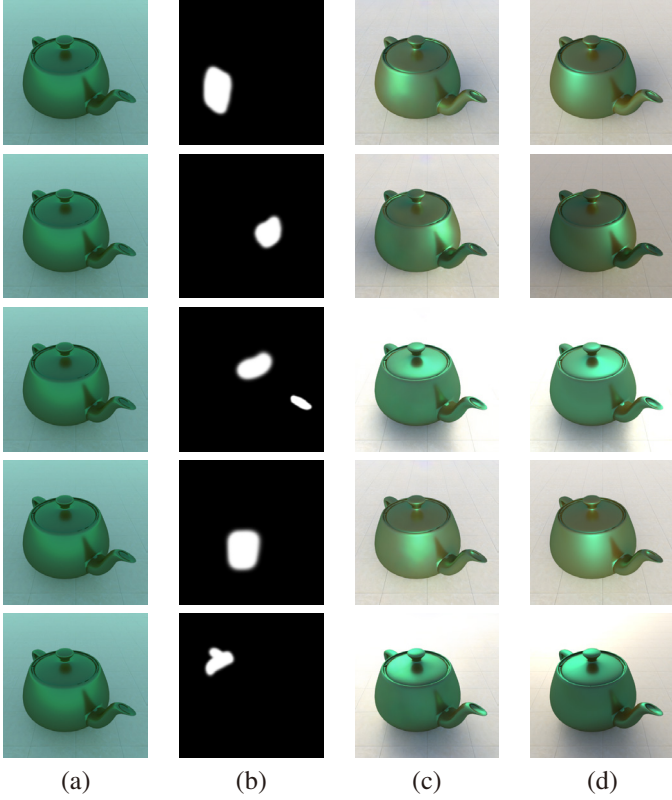


Figure 4: Result of our framework with the Utah teapot model. (a) base image, (b) user annotated specular highlight map, (c) **our** neural relighted image, (d) rendered image obtained from a conventional 3D graphics renderer using **our** predicted light position.

평가 위의 방법으로 계산된 반사 하이라이트 맵을 입력으로 조명 위치와 재조명된 이미지가 실제 조명의 위치, 이미지와 얼마나 동일한지 측정하였다. 실제 조명과 추정된 조명의 위치는 유클리디언 거리(Euclidean distance)로 차이를 측정하고, 실제 조명의 위치를 포함하는 장면의 바운딩 박스(bounding box)의 대각선 길이에 대한 비율을 계산하였다. 실제 영상과 재조명된 영상의 차이는 PSNR로 측정하였다.

Fig. 6는 제안하는 방법으로 테스트 데이터셋에 대해서 조명의 위치를 추정하고 재조명한 결과를 보여준다. Fig. 6(e)는 실제 영상 Fig. 6(c)과 재조명한 영상 Fig. 6(d)의 차이를 나타낸 것으로, 제안하는 방법이 실제 영상과 유사하게 재조명한 것을 확인할 수 있다. Table 2에서 추정한 조명의 위치의 차이는 평균 0.11이고 PSNR은 평균 28.89 dB인 것을 확인할 수 있다. 이 때 실제 조명 위치를 포함한 바운딩 박스를 설정해 이에 대한 대각선 길이를 기준으로 추정한 조명 위치를 정규화하여 오차를 측정하였다.

Fig. 7은 조명의 위치에 따른 PSNR 수치의 변화를 시각화한 결과이다. 실험을 위해 100개의 실제 조명 위치를 전역좌표계를 기준으로 방위각(azimuth)과 고도각(elevation)에 따라 배치한 후, 각 위치에서 실제 영상과 재조명된 영상의 차이를 PSNR로 측정하였다. Fig. 7에서 상대적으로 큰 원은 높은 PSNR을 나타내고, 작은 원은 낮은 PSNR을 나타낸다. Fig. 7을 통해 고도각이 PSNR에 영향을 미치고 있으며, 특히 바닥 평면 대비 조명의 고도가

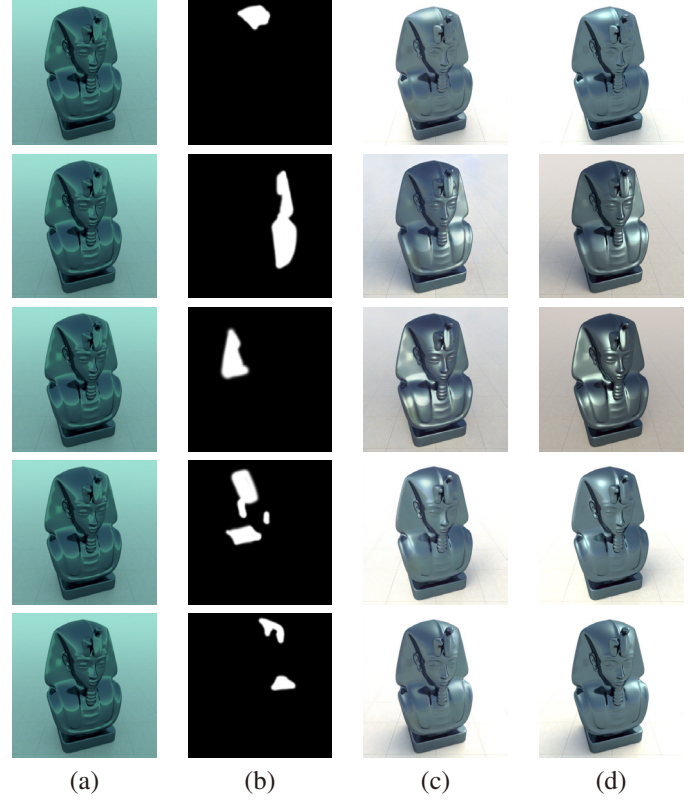


Figure 5: Result of our framework with the Pharaoh model (courtesy of Free3D.com). (a) base image, (b) user annotated specular highlight map, (c) **our** neural relighted image, (d) rendered image obtained from a conventional 3D graphics renderer using **our** predicted light position.

낮을 수록 결과 영상의 PSNR이 높아지는 것을 확인할 수 있다.

Fig. 8은 조명의 실제 위치에 대한 방위각과 고도각을 각각 [0.2, 1.2, 2.2, 3.2, 4.2, 5.2]과 [0.2, 0.7, 1.2, 1.7, 2.2, 2.7] 라디안(radian)으로 변화시키며 얻은 재조명 결과를 보여주고 있다. Fig. 9은 Fig. 8을 만들어내기 위한 실제 영상과 반사 하이라이트 맵을 보여준다. Fig. 9의 2번째, 3번째, 4번째 행에서 보이는 것처럼 바닥 평면 대비 실제 조명의 고도가 높으면 실제 영상이 기저 영상에 비해 밝아지기 때문에 반사 하이라이트 맵이 배경을 포함한 영상 전체에 대해 생성된다. 이 경우 실제 반사 하이라이트 맵이 거의 대부분의 픽셀에서 형성되기 때문에 제안하는 뉴럴 재조명이 이를 만족하도록 최적화를 진행하게 되어 최종 재조명 결과의 PSNR은 상대적으로 낮아지게 된다. 반면에 실제 조명의 고도가 낮으면 실제 영상과 기저 영상의 밝기 차이가 일부의 픽셀에서만 드러나기 때문에 제안하는 뉴럴 재조명 기법이 실제 반사 하이라이트 맵의 특정 부분에 대해서만 최적화를 진행하게 되어 재조명 결과 영상의 PSNR이 상대적으로 높아지게 된다.

5. 결론 및 향후 연구

본 논문에서는 사용자가 지정한 반사 하이라이트 맵을 가이드 영상으로 하는 신경망 기반 재조명 기법을 제안하였다. 기저 영상

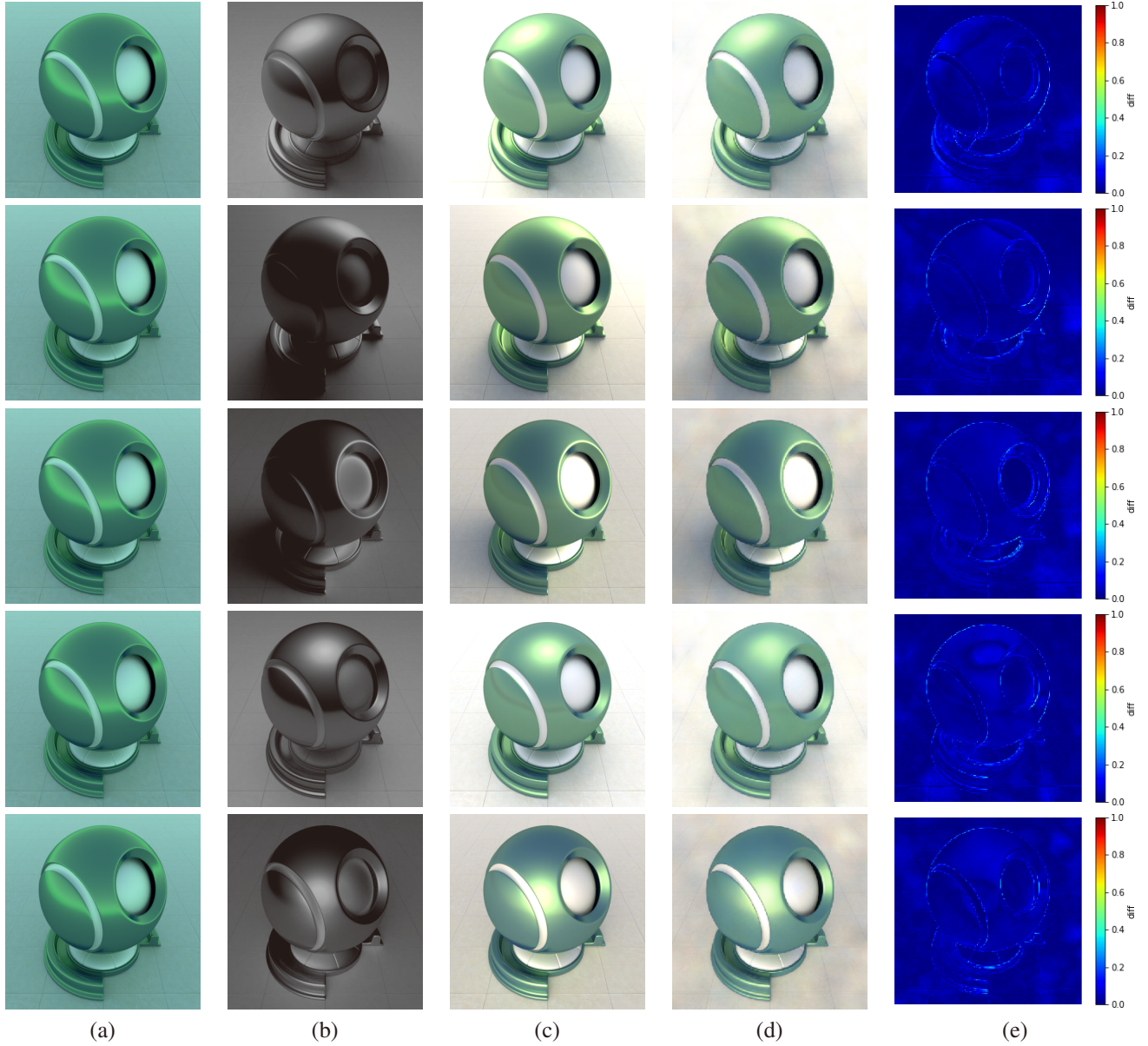


Figure 6: Result of our framework with GT. (a) base image, (b) specular highlight map, (c) GT image obtained from a conventional 3D graphics renderer, (d) **our** neural relighted image, (e) difference between **our** neural relighted image and GT image.

과 렌더링 영상의 차이가 사용자가 제공한 반사 하이라이트 맵과 유사하도록 역전파에 의해 광원의 위치와 관련된 재조명 이미지를 동시 최적화하는 방식으로 아티스트가 선호하는 이차원 화면 공간 인터페이스를 제공하면서도 삼차원 조명의 위치를 명시적으로 추론할 수 있는 장점이 있다.

향후 본 논문에서 제안하는 기법을 아티스트의 편의성을 유지 하면서 보다 다양한 삼차원 장면에 대한 결과를 예측할 수 있도록 다양한 방식으로 확장할 계획이다. 현재 하나의 모델과 시점에 대해서만 작동하는 프레임워크를 다양한 모델과 시점에서도 동작이 가능하도록 확장할 계획이다. 다음으로 여러 개의 조명을 추정하는 연구와 사용자가 반사 하이라이트뿐만 아니라 그림자를 지정하면 이를 만족하는 조명 조건을 추정하는 연구를 진행 중이다.

감사의 글

본 논문은 교육부와 과학기술정보통신부의 재원으로 한국연구재단의 지원을 받은 기초연구사업(2017R1D1A1B03034907), 선도연구센터지원사업(2015R1A5A7037615)의 일환으로 수행된 연구임.

References

- [1] M. Pharr, W. Jakob, and G. Humphreys, *Physically Based Rendering: From Theory to Implementation*, 3rd ed. Morgan Kaufmann, 2017.

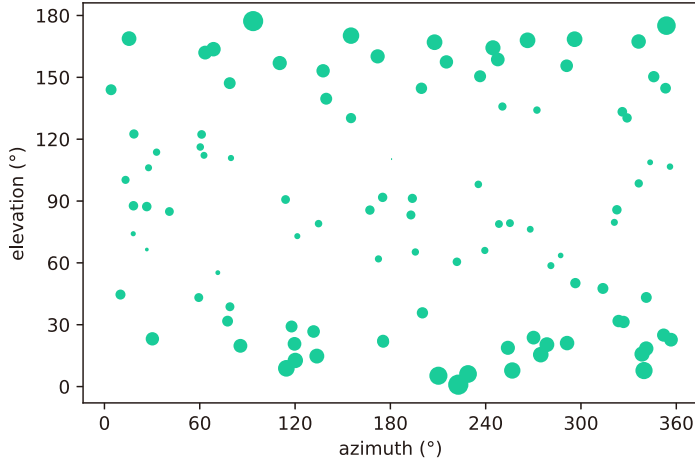


Figure 7: Visualization of PSNR values for different light directions. The position of each circle represents the azimuth and elevation of the light, and the size of each circle represents the relative PSNR value (dB). The larger circle size represents the higher PSNR value (dB), and vice versa.

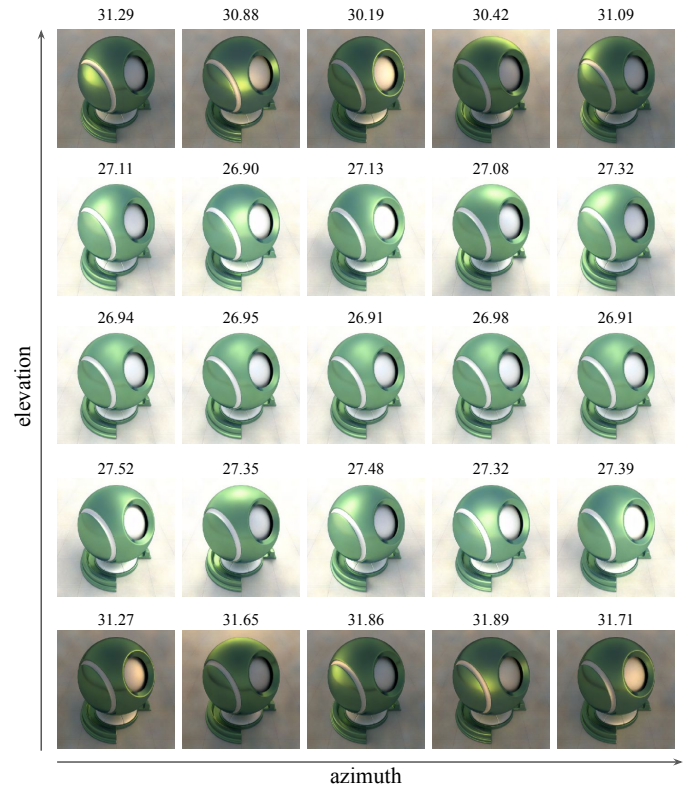


Figure 8: Result of our framework for different light directions. The values above the images represent PSNR values (dB).

- [2] A. Tewari, O. Fried, J. Thies, V. Sitzmann, S. Lombardi, K. Sunkavalli, R. Martin-Brualla, T. Simon, J. Saragih, M. Nießner, R. Pandey, S. Fanello, G. Wetzstein, J.-Y. Zhu, C. Theobalt, M. Agrawala, E. Shechtman, D. B. Goldman, and M. Zollhöfer, “State of the Art on Neural Rendering,” *Eurographics STAR*, vol. 39, no. 2, pp. 701–727, 2020.
- [3] A. KT, P. Sakurikar, S. Saini, and P. J. Narayanan, “A Flexible Neural Renderer for Material Visualization,” in *SIGGRAPH Asia 2019 Technical Briefs*, 2019, pp. 83–86.
- [4] J. Thies, M. Zollhöfer, and M. Nießner, “Deferred Neural Rendering: Image Synthesis using Neural Textures,” *ACM Trans. Graphics*, vol. 38, no. 4, pp. 66:1–66:12, 2019.
- [5] Z. Chen, A. Chen, G. Zhang, C. Wang, Y. Ji, K. N. Kiriakos, and J. Yu, “A Neural Rendering Framework for Free-Viewpoint Relighting,” in *Proc. CVPR*, 2020, pp. 5599–5610.
- [6] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde, “Learning to Predict Indoor Illumination from a Single Image,” *ACM Trans. Graphics*, vol. 36, no. 6, pp. 176:1–176:14, 2017.
- [7] M.-A. Gardner, Y. Hold-Geoffroy, K. Sunkavalli, C. Gagné, and J.-F. Lalonde, “Deep Parametric Indoor Lighting Estimation,” in *Proc. ICCV*, 2019, pp. 7175–7183.
- [8] T. Y. Wang, T. Ritschel, and N. J. Mitra, “Joint Material and Illumination Estimation from Photo Sets in the Wild,” in *Proc. 3DV*, 2018, pp. 22–31.
- [9] H. Zhou, S. Hadap, K. Sunkavalli, and D. W. Jacobs, “Deep Single-Image Portrait Relighting,” in *Proc. ICCV*, 2019, pp. 7194–7202.
- [10] P. Kán and H. Kafumann, “DeepLight: light source estimation for augmented reality using deep learning,” *The Visual Computer*, vol. 35, no. 6-8, pp. 873–883, 2019.
- [11] P.-P. J. Sloan, W. Martin, A. Gooch, and B. Gooch, “The Lit Sphere: A Model for Capturing NPR Shading from Art,” in *Proc. Graphics Interface*, 2001, pp. 143–150.
- [12] A. Hertzmann, C. E. Jacobs, N. Oliver, B. Curless, and D. H. Salesin, “Image Analogies,” in *Proc. ACM SIGGRAPH 2001*, 2001, pp. 327–340.
- [13] P. Bénard, F. Cole, M. Kass, I. Mordatch, J. Hegarty, M. S. Senn, K. Fleischer, D. Pesare, and K. Breeden, “Stylizing Animation By Example,” *ACM Trans. Graphics*, vol. 32, no. 4, pp. 119:1–119:12, 2013.
- [14] J. Fišer, O. Jamriška, M. Lukáč, E. Shechtman, P. Asente, J. Lu, and D. Sýkora, “StyLit: Illumination-Guided Example-Based Stylization of 3D Renderings,” *ACM Trans. Graphics*, vol. 35, no. 4, pp. 92:1–92:11, 2016.



Figure 9: GT images (left 5 columns) and corresponding specular highlight maps (right 5 columns) for different light directions.

- [15] L. Zhang, E. Simo-Serra, Y. Ji, and C. Liu, “Generating Digital Painting Lighting Effects via RGB-space Geometry,” *ACM Trans. Graphics*, vol. 39, no. 2, pp. 13:1–13:13, 2020.
- [16] L. Hosek and A. Wilkie, “An Analytic Model for Full Spectral Sky-Dome Radiance,” *ACM Trans. Graphics*, vol. 31, no. 4, pp. 95:1–95:9, 2012.
- [17] “Blender,” <https://www.blender.org/>.

〈 저 자 소 개 〉



이연경

- 2020년 2월 국민대학교 시각디자인학과 학사
- 2020년 3월~현재 국민대학교 컴퓨터공학과 석사과정
- 관심분야: 컴퓨터 그래픽스, 딥러닝
- <https://orcid.org/0000-0002-7563-0328>



고 현 성

- 2019년 3월~현재 국민대학교 소프트웨어학부 학사과정
- 관심분야: 컴퓨터 그래픽스, 게임 엔진
- <https://orcid.org/0000-0001-7358-464X>



이 진 우

- 2011년 2월 국민대학교 컴퓨터공학과 학사
- 2013년 2월 국민대학교 컴퓨터공학과 석사
- 2013년 3월~현재 국민대학교 컴퓨터공학과 박사과정
- 관심분야: 컴퓨터 그래픽스, 게임 엔진, 증강 현실, 딥러닝
- <https://orcid.org/0000-0002-9171-2016>



김 준 호

- 1988년 2월 포항공과대학교 컴퓨터공학과 학사
- 2000년 2월 포항공과대학교 컴퓨터공학과 석사
- 2005년 2월 포항공과대학교 컴퓨터공학과 박사
- 2009년 3월~현재 국민대학교 소프트웨어학부 부교수
- 관심분야: 컴퓨터 그래픽스, 3차원 모델, 증강 현실, 딥러닝
- <https://orcid.org/0000-0003-0473-7053>