

옷감 이미지 렌더링을 위한 Pix2Pix 기반의 Normal map 생성

남현길, 박종일¹

한양대학교

skagusrlf@hanyang.ac.kr, jipark@hanyang.ac.kr

Normal map generation based on Pix2Pix for rendering fabric image

Hyeongil Nam Jong-Il Park

Hanyang University

요 약

본 논문은 단일의 옷감 이미지로 가상의 그래픽 렌더링을 위해 Pix2Pix 방법을 이용하여 Normal map 을 생성하는 방법을 제시한다. 구체적으로 단일의 이미지를 이용해서 Normal map 를 생성하기 위해, Color image 와 Normal map 쌍의 training dataset 을 Pix2Pix 방법을 이용해서 학습시킨다 또한, test dataset 의 Color image 를 입력으로 넣어 생성된 Normal map 결과를 확인한다. 그리고 선행연구에서 사용되어오던 U-Net 방식의 방법과 본 논문에서 사용한 Pix2Pix 를 이용한 Normal map 생성 결과를 SSIM(Structural Similarity Index)으로 비교 평가한다. 또한, 생성된 Normal map 을 렌더링하고자 하는 가상 객체의 사이즈에 맞게 사이즈를 조정하여 OpenGL 로 렌더링한 결과를 확인한다. 본 논문을 통해서 단일의 패턴 이미지를 Pix2Pix 로 생성한 Normal map 으로 옷감의 디테일을 사실감 있게 표현할 수 있음을 확인할 수 있었다.

1. 서론

빛에 반사되는 옷감의 형태를 보다 사실적으로 보여주기 위해서 Color 이미지와 함께 Normal map 을 사용하는 경우가 많다. 특히나, 실제 표면에 굴곡이 있는 옷감의 경우에는 Normal map 을 활용하는 빛의 반사 특성을 더욱 효과적으로 표현할 수 있다. Normal map 을 생성하기 위해서 그래픽 디자이너가 Photoshop 이나 다른 응용 프로그램을 이용하여 개별적인 작업이 필요했다. 최근의 연구에서는 딥 러닝을 이용하여 단일 이미지의 Normal map 을 생성하는 연구가 활발하게 이루어지고 있다[1][2]. 선행연구에서는 Autoencoding 방식의 U-Net 을 이용하여 단일의 이미지를 Normal map 으로 변환하였다[1][2]. U-Net 의 경우에는 encoding 과 decoding 과정을 거치면서 발생하는 정보손실을 보완하기 위한 방법을 위주로 되었다[5].

그러나 이와 달리 본 논문에서는 영상 변환 및 생성을 위해서 사용되는 GAN(Generative Adversarial Networks) 기반의 Pix2Pix 방식으로 Normal map 을 생성하고자 한다.

본 논문에서는 다음과 같은 구성으로 연구를 진행하였다. 2 절에서는 Color Image, Normal map 쌍으로 구성된 231 개의 Dataset 을 Training set 과 Test set 으로 구분하여 Pix2Pix 로 Training Set 을 학습시킨 방법에 대해 설명한다. 그리고 3 절에서는 학습된 결과를 Color Image 를 이용해 생성된 Normal map 과 선행연구의 방식으로 생성된 Normal map 을 각각 Ground Truth Normal map 과 비교 평가한다. 비교 평가를 위해서 SSIM(Structural Similarity Index)를 이용하여 Ground Truth 에 해당하는 Normal map 과 각 방식으로 생성된 Normal map 의 구조적 유사도를 판단한다. 또한, 4 절에서는 각 방식으로 생성된 Normal map 을 OpenGL 로 Rendering 된 결과를

¹ 교신저자

육안으로 비교 평가한다. 본 논문을 통해서 패턴이 있는 옷감 텍스처를 보다 사실감 있게 렌더링하기 위해 Pix2Pix 방법으로 Normal map 을 생성할 수 있음을 확인하였다.

2. Pix2Pix 기반의 Normal map 생성

기존의 선행연구는 Color image 를 Auto encoding 방식의 U-Net 으로 Normal map 을 생성하는 방식이 주목을 받았다[1]. Auto encoding 의 압축과정에서 발생하는 정보의 손실을 보완하기 위해서, Down sampling 에서 발생하는 특징들을 Up sampling 시에 함께 반영하였다[1]. 또한 Global feature 를 전체 학습과정에 반영하여 정보의 손실을 보완하는 등의 방법을 사용한 연구도 있었다[2]. 본 논문에서는 이와 달리, GAN 기반의 생성 모델인 Pix2Pix 를 통해서 Normal map 을 생성하는 방법을 제안한다. Pix2Pix 의 경우에는 Generator(G)와 Discriminator(D)를 사용하는 기본적인 GAN model 의 특징을 모두 가지고 있다[3]. Pix2Pix 로 Normal map 을 학습시키기 위한 기본적인 구조는 다음 그림 1 과 같이 구성되었다. 학습을 위해서 231 개의 Color image(256x256)와 학습의 목표 영상에 해당하는 Normal map(256x256)을 이용하여, Training Set(80%)와 Test Set(20%)으로 나누었다.

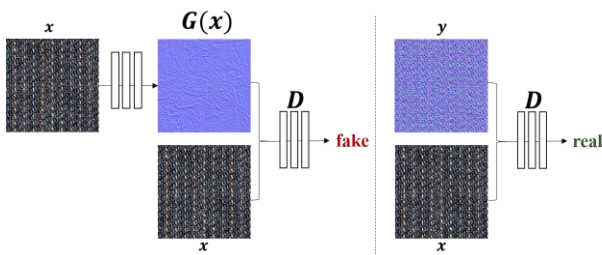


그림 1. Normal map 생성을 위한 Pix2Pix 네트워크의 구조

학습을 위한 Dataset 의 경우에는 Pixar Textures 를 사용하였다[4]. Pix2Pix 의 내부적 구조를 살펴보면, Generator 의 경우에는 선행연구에서 차용되었던 U-Net 을 동일하게 사용한다[5]. 또한 Loss Function 도 마찬가지로 L1(reconstruction loss)과 cGAN(adversarial loss)을 함께 사용하여 U-Net 에서의 성능을 선행연구에 비해서 끌어올릴 수 있었다[선행 5]. Generator 에서는 Ground Truth 영상과 생성된 영상 사이의 유클리드 거리를 최소화하는 방향인 L1 loss 를 사용하여 저주파 영역에 집중을 하고, PatchGAN 을 사용하는 Discriminator 의 경우에는 patch 단위로 진짜와 가짜를 판단하고 이를 평균을 취하게 되어 고주파에 집중하게 된다[5]. 이를 통해서 더욱 선명한 Normal map 을 생성할 수 있게 되었다. 또한 이후 비교 평가를 위해 동일한 데이터셋에 대해 선행연구의 U-Net 방식으로 추가 학습을 수행하였다[1].

3. SSIM 비교

Pix2Pix 방법을 이용하여 생성한 결과를 선행연구에서 사용된 U-Net 방법의 결과를 비교 평가가 필요하다. 이를 위해서 원본 영상과 각 방법의 SSIM(Structural Similarity Index)을 평가 지표로 사용하였다. SSIM 은 구조적 유사 지수를 의미한다[6]. 보통 영상의 압축 및 변환에 의해서 발생하는 왜곡에 대해 원본 영상에 대한 유사도를 측정하는 방법으로서 MSE(Mean Square Error), PSNR(Peak Signal-to-Noise Ratio) 방법보다 더 정확한 비교가 가능하다[6]. SSIM 의 경우에는 구조 정보의 왜곡 정도가 사람이 지각하는 품질에 영향을 미친다는 가정하에서 원본 이미지 x 와 왜곡 이미지 y 의 밝기, 콘트라스트, 구조를 비교한다[6]. 이를 반영한 수식은 다음과 같다(그림 2).

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(2\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

μ_x : x의평균, μ_y : y의평균, σ_x^2 : x의분산, σ_y^2 : y의 분산,
 σ_{xy} : x,y의공분산, $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$: weak denominator
 를 안정시키기 위한 두 변수, L: 픽셀 값의 동적인 범위(전형적 으로는 $2^{\#bits \text{ per pixel}} - 1$), $k_1 = 0.01, k_2 = 0.03$ 으로 디폴트 값

그림 2. SSIM(Structural Similarity Index)의 식과 설명

Ground Truth 와 Pix2Pix 방식으로 생성한 Normal map 의 SSIM_{Pix2Pix} 와 Ground Truth 와 U-Net 방식으로 생성한 Normal map 의 SSIM_{U-Net} 을 각각 구한다. Test dataset(46 개 samples)에 대해 그 차이를 구한 결과가 그림 3 과 같다.

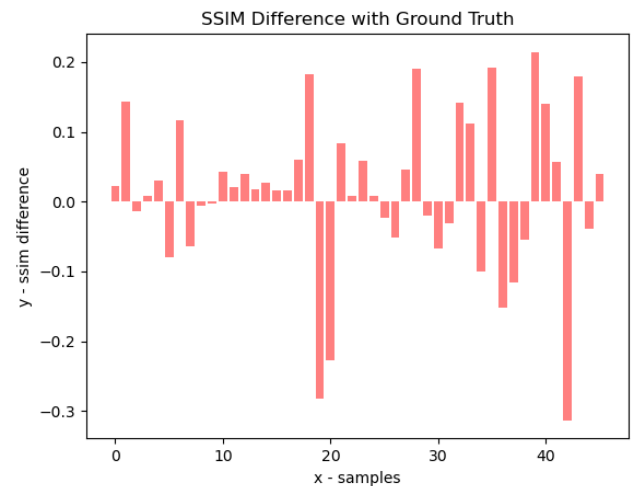


그림 3. Ground Truth Normal map 과 Pix2Pix 방법의 Normal map 사이의 SSIM_{Pix2Pix} 과 Ground Truth Normal map 과 U-Net 방법의 Normal map 의 SSIM_{U-Net} 의 차이

Test dataset 의 samples 중에서 60%이상의 sample 에서 Pix2Pix 를 이용하여 생성한 Normal map 과 Ground Truth 의 Normal map 사이에서 구한 SSIM 이 더 높게 나타났다.

4. Normal Map 및 렌더링 결과 비교

Ground Truth Normal map 과 본 논문에서 사용한 방법인 Pix2Pix 를 이용하여 생성한 Normal map, 그리고 U-Net 을 사용한 경우 예측된 Normal map 을 그림 4 와 같이 육안으로 비교한다. U-Net 의 경우에는 압축과 해제 과정의 학습을 통해서, detail 한 부분의 feature 들이 많이 사라져버린 것을 확인할 수 있었다. 반면에 pix2pix 방법을 사용한 본 논문의 실험에서는 feature 를 U-Net 에 비해 비교적 feature 들을 Ground Truth 와 유사하게 생성할 수 있었다. 그러나 Ground Truth 에서 feature 가 없어야 할 부분에도 feature 가 생성되는 경우도 발생하였다.

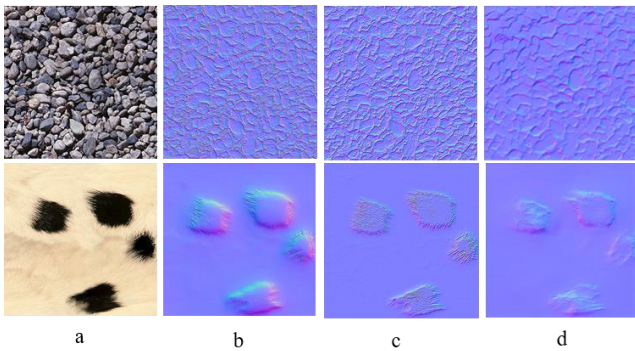


그림 4. [위-big peddles, 아래-panda fur]: (a)Ground Truth Color image, (b)Ground Truth Normal map, (c)Pix2Pix 의 Normal map, (d)U-Net 의 Normal map

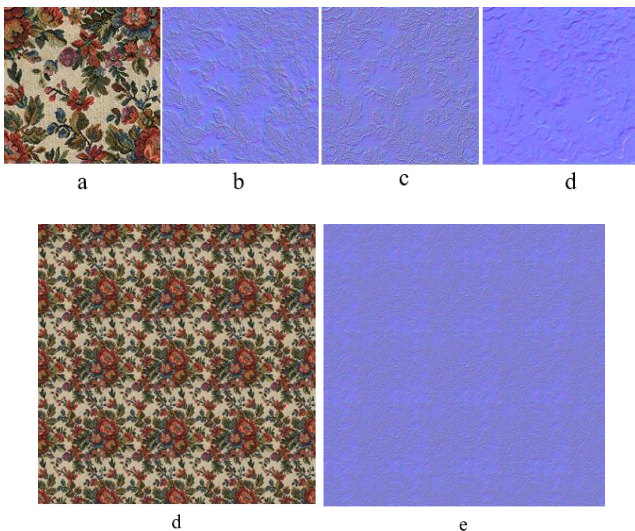


그림 5. Flower pattern (a)Ground Truth Color image (256x256), (b)Ground Truth Normal map, (c)Pix2Pix 의 Normal map, (d)U-Net 의 Normal map, (e)렌더링을 위한 Color image size 변경(1024x1024), (e) Pix2Pix 의 Normal map size 변경(1024x1024) 예시

실제로 Flower pattern 의 옷감 이미지를 그래픽으로 렌더링 된 결과를 확인하기 위해, 그림 5 와 같이 렌더링을 위해서 Color image 와 Normal map 의 size 를 256x256 에서 1024x1024 로 변경하였다. 그리고 그림 6 과 그림 7 에서 만들어진 Ground Truth Normal map 과 생성된 방법들의 Normal map 을 이용하여 실제 Rendering 한 결과를 비교한다.



그림 6. 세 가지 이미지는 동일한 Color image 를 사용하였고, (a) Ground Truth Normal map, (b) Pix2Pix 의 Normal map, (c) U-Net 의 Normal map 으로 가상의 바지에 렌더링한 결과 ((x,y,z = (0,0,30))위치의 점 광원, 100W, 주변광(1.0))

그림 6 은 주변광이 밝은 상황(1.0)에서의 렌더링 결과를 비교한다. 이 때에는 Ground Truth 대비 생성된 normal map 효과의 차이를 확인하기 어렵다. 그러나 그림 7 과 같이 주변광이 어두운 경우(0.3)에는 Pix2Pix 의 normal map 과 U-Net 의 Normal map 으로 각 Rendering 된 결과에서 빛의 반사특성의 차이를 더 극명하게 확인할 수 있었다. 붉은색 네모 박스처럼 그림 7 의 (b)Pix2Pix 로 만들어진 Normal map 을 렌더링했을 때가 (c)U-Net 으로 생성된 Normal map 으로 렌더링된 바지에 비해 빛의 반사 효과를 더욱 잘 표현하였다.

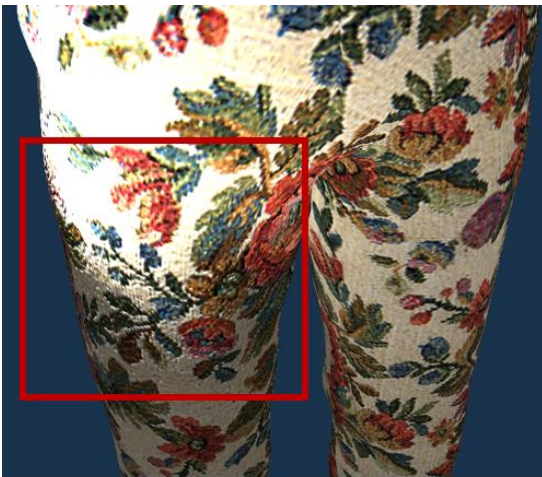
5. 결론

본 논문은 단일의 옷감 이미지로 Pix2Pix 를 이용하여 Normal map 을 생성하는 방법을 제안하였다. Ground Truth 에 해당하는 Color image 와 Normal map 쌍을 Training dataset 으로 학습시키고, 그 결과를 Test dataset 을 통해서 확인할 수 있었다. 사용된 방법의 효과를 판단하기 위해 선행연구에 사용된 U-Net 방식의 Normal map 을 동일한 dataset 에 대해 적용하였다. 그리고 Ground Truth Normal map, Pix2Pix 의 Normal map 사이의 SSIM(Structural Similarity Index)과 Ground Truth Normal map, U-Net 의 Normal map 사이의 SSIM 을 계산하여 비교하였다. 그 결과 60%의 sample 에서 Pix2Pix 방식이 Ground Truth 와 구조적으로 더 유사하다고 평가하였다. 또한 실제 가상의 객체에 각 방법으로 생성된 Normal map 을 적용하여 Rendering 된 결과를 육안으로도 확인하여 더욱 반사특성을 잘 표현하는 것을

확인할 수 있었다.



a



b



c

그림 7. (a) Ground Truth Normal map, (b) Pix2Pix 의 Normal map, (c) U-Net 의 Normal map 으로 가상의 바지에 랜더링한 결과 ((x,y,z = (-10,0,40))위치의 점 광원, 100W, 주변광(0.5))

감사의 글

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 지원을 받아 수행된 연구임 (No.2017-0-01849, 실내의 임의공간 실시간 영상 합성을 위한 핵심 원천기술 및 개발툴킷 개발)

참고문헌

- [1] GAO, DUAN & Li, Xiao & Dong, Yue & Peers, Pieter & Xu, Kun & Tong, Xin. Deep inverse rendering for high-resolution SVBRDF estimation from an arbitrary number of images. ACM Transactions on Graphics. 38. 1-15. 10.1145/3306346.3323042, 2019
- [2] Deschaintre, Valentin et al. "Single-image SVBRDF capture with a rendering-aware deep network." ACM Transactions on Graphics (TOG) 37 2018: 1 - 15.
- [3] Isola, Phillip et al. "Image-to-Image Translation with Conditional Adversarial Networks." 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016: 5967-5976.
- [4] Pixar One Twenty Eight by Pixar Animation Studios, <https://renderman.pixar.com/>
- [5] Ronneberger, Olaf & Fischer, Philipp & Brox, Thomas. U-Net: Convolutional Networks for Biomedical Image Segmentation. 2015.
- [6] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. IEEE transactions on image processing, 13(4):600, 2004.