

FACSIMILE:

Fast and Accurate Scans From an Image in Less Than a Second

번역본(의역 및 오류 주의)

<요약>

근래의 체형 추론 방법들은 디테일이 떨어지고, 많은 이미지가 필요하다. 그렇기에 우리는 FAX(FACSIMILE)를 제안한다. 이것은 한장의 사진으로 부터 신체의 디테일을 추론하는 방법론이고, 이로 인해 사람을 가상에서 재현하는 것에 대해 문턱을 낮춘다. 역설적이게도, FAX의 메인 loss는 normal에 대한 픽셀당 로스이다, 깊이의 픽셀당 로스가 아니다. 이것은 깊이 감독(supervision) 없이도 신체의 기하학적 디테일을 만들게하기 가능하게 만든다. 우리는 우리의 접근방법을 정성적 정량적으로 평가하고, 기존 SOTA모델과 비교하였다.

<1. 서론>

의학, 게임, 쇼핑의 많은 활용에도 불구하고 고해상도의 신체 작업물(body capture)은 대중적으로 채택되지 않아왔다. 전통적인 방식의 고해상도의 신체 추론은 사용하기 어려운(difficult to deploy), 비싼 시스템이 필요하다. 보다 사용가능한 kinetic과 같은 RGB-D 센서들은 이러한 문제들을 극복 해왔으나, RGB카메라보다 대중적이지 않다(widespread). 반면에, 한장의 사진으로 신체를 추론하는 근래의 시스템은, 디테일이 부족하다. 우리의 작업은 쉽게 획득할 수 있는 이미지와 풍부하고 자세하며 자세를 바꿀 수 있는 아바타 사이의 격차를 줄이는 데 도움이 되도록 설계되었다.

단일 이미지로 부터 모양(shape)을 복원하는 것을 목표로 삼은 시스템은 몸통(intermediate body) 재현을 복원하는 것에 있어서 칭찬할만한(laudable) 작업이다. 이는 복셀기반 재현, 인공적인-시각 생성 시스템, 또는 교차모달 신경망을 포함한다. 그러나 불가피하게도, 그들의 캡처의 가치는 (fidelity) 표현의 세분성으로 인해 제한된다.

이러한 표현의 힘의 부족을 다루기 위해, 우리는 기하학 추론에 있어서 근래의 이미지 대 이미지 변환 기술을 적용하였다. 좀더 구체적으로, 우리는 이미지내의 모든 픽셀에 대해 깊이 추론을 하고싶다. 그러나 이것은 새로운 문제를 야기한다: 이미지 변환을 통한 단순한 깊이 추론은 노이즈와 사용할수 없는 표면을 만들어낸다. 이것은 이미지 대 이미지 변환으로 깊이를 추정할 때 깊이 에 대한 직접적인 loss가 그럴듯한 표면을 제공하지 못한다는 것을 가르쳐준다.

이것에 대한 해결책은 Horn의 Shpe From Shading 문헌으로 거슬러 갈 수 있다. 표면 법선은 표면과 표면의 모양 사이의 관계를 정의하는 데 중요한 역할을 한다(normal map에 대한 기초 지식이 필요함). 얼굴 영역의 재구성에 초점을 맞춘 작업은 깊이 손실이 법선에 대한 추가 손실로부터

이익을 얻을 수 있음을 보여주었다. 우리는 법선에 대한 손실만으로도 규모에 맞게 고품질 깊이 맵을 재구성하기에 충분할 수 있으며 이는 구면에서 멀리 떨어진 관절형 객체에 적용된다는 것을 보여주는 통찰력을 넘어섰다.

단일의 깊이맵은 전체 아바타와는 거리가 멀기 때문에, 우리는 정면과 후면의 기하와, 알베도를 추론하도록 시스템을 확장하였다(albedo의 기초지식이 필요함). 우리는 사람의 뒷면을 추론하도록 네트워크를 학습시킴으로서 픽셀당 두 개의 값을 얻는다는 아이디어를 차용한다. 이전의 연구와는 다르게, 텍스처에 제한을 두지 않고, 후면 깊이와 법선도 추정한다. 현재의 세부적인 방법들은 일반적으로 실행하는 데 몇 분이 걸리지만 기하와 텍스처를 포함하는 거의 완전한 스캔을 1초 이내에 계산한다. 본 논문에서 우리는 어느정도의 변수 통제(cooperative subject)를 가정하고 정보 캡처를 최대화하는 특정 유형의 이미지(전면 팔을 아래로 하는 자세, 최소한의 옷차림)에 초점을 맞추지만 이 방법이 다른 경우에도 적용될 수 있다고 믿고 향후 작업에서 계속 조사할 것이다.

우리는 세 가지 기여가 있다. 첫째, 우리는 단일 이미지에서 전체 스캔을 계산하는데, 이는 세부 스캔을 생성하는 현재 방법보다 훨씬 더 빠르다. 다른 방법도 의복을 재현하지만 우리 방법은 훨씬 더 많은 세부 사항을 추출한다. 둘째, 이러한 스캔이 게임, 이미지 측정 및 가상 영상회의와 같은 응용 프로그램에 유용할 수 있는 추가 시간(10초 미만)이 거의 없이 변형 가능한 세부 아바타로 변환될 수 있는 방법을 보여준다. 마지막으로 우리는 우리 방법의 효율성을 최첨단 다중 이미지 방법과 정량적으로 비교하고 정성적 및 정량적 ablation study를 수행하여 설명한다.

<2. 관련 연구>

Geometry estimation from a single-photo는 적어도 50년 동안 연구 주제였다. shape from shading과 같은 고전적인 방법은 음영 이미지를 가져와 기본 기하를 생성한다. 이 문제에 대한 최신 솔루션은 계산적으로 효율적이고 직관적일 수 있지만, 데이터에 적용된 조명 및 분포 모델의 한계로 인해 실제 데이터에서 피할 수 없는 입력 노이즈가 있을 때 부서지기 쉽다. 딥 러닝 기반 방법은 자율 주행을 위한 실외 깊이 재구성 및 실내 기하학 재구성에서 이러한 취성(brittleness/부서지기 쉬움)을 줄이는 데 인상적인 결과를 달성하였다.

Single-photo body estimation은 일반적으로 고정된 중간 표현을 통해 병목 현상이 발생하며, 이는 조각별 모델링을 가능하게 하지만 궁극적으로 달성 가능한 세부 정보의 양을 제한한다. 일부 방법은 분할된 이미지를 통해 병목 현상을 발생시키고, 다른 방법은 추정된 키포인트 위치를 통해, 일부는 둘 다를 통해 병목 현상을 일으킨다. 이러한 모든 방법은 조밀한 표면 재구성을 허용하기에는 너무 많은 모호성을 허용한다. 최근 방법은 인코더 디코더 표현을 이미지에 직접 사용하여 이러한 제한을 피한다. 그들은 다방면(in the wild)에 대해 놀라운 견고성을 달성하지만 세부적인 모양과 자세를 복구하는 데 어려움을 겪는다. SURREAL에 대한 작업은 깊이를 직접 추정하지만 세부 사항은 조악하다. SiCloPe 시스템은 우리 시스템보다 더 큰 의복 변화를 허용하지만 기하학적 세부 사항은 중간 실루엣의 사용으로 인해 제한된다. 이 작업의 공로로 (Human Shape from

Silhouettes Using Generative HKS Descriptors and Cross-Modal Neural Networks)을 제외한 모든 것은 포즈 변형에 대한 허용 오차로 "in the wild" 신체를 캡처하도록 설계되었지만 우리의 목표는 제한된 포즈에서 상세한 아바타를 캡처하는 것이다.

Single-photo face estimation은 신체 추정에 유용한 통찰력을 제공한다. Blanz와 Vetter의 초기 작업은 혁신적이었지만 세부 사항이 부족하고 다방면의(in the wild) 견고성에 문제가 있었습니다. 견고성은 데이터 기반 모델에 의해 해결되었다. 세부 사항은 먼저 음영의 모양으로 해결된 다음, 딥러닝으로 해결되었다. Zollhofer et al의 최근 설문조사는 보다 구체적이다. FAX는 [Unrestricted Facial Geometry Reconstruction Using Image-to-Image Translation, High-Fidelity 3D Digital Human Head Creation from RGB-D Selfies]와 구체적으로 주제를 공유하는데, 여기서 Isola et al의 이미지-이미지 변환 아키텍처가 세부적인 얼굴 기하학 추정에 성공적으로 적용되었다. 우리의 초점은 단일 RGB 이미지에서 아바타 기하 추정에 있다. 여러 이미지의 신체 추정에 대한 보다 일반적인 검토를 위해 독자는 Alldieck et al 및 Bogo et al에서 제공된 이전 작업의 우수한 요약물 검토하는 것이 좋다.

<3. Method>

우리의 목적은 단일 RGB 이미지로 세밀한 3차원 스캔을 추정하는 것이다. 우리는 이것을 이미지 대 이미지 변환을 통해 다룰 것이고, 이미지를 깊이 그리고 알베도 값으로 변환 할 것이다. 좀 더 구체적으로, 우리는 이러한 추정을 신체의 앞과 뒤를 할 것이다. 깊이 이미지는 3D 표면을 만들기 위해 간단하게 삼각 측량할 수 있는 정점의 규칙적인 격자를 형성한다. 우리는 깊이 추정 아키텍처를 3.2절에서 더 자세히 설명하지만 훈련 프로토콜이 [46]의 이전 작업과 매우 유사하기 때문에 먼저 3.1절에서 알베도 추정에 중점을 둔다. 마지막으로 섹션 3.3에서 완전하고 재배포 가능하며 재구성 가능한 아바타를 얻는 방법을 설명한다.

<3.1 Albedo estimation>

우리의 아키텍처는 이미지 대 이미지 변환의 일환인 cGAN에 기반하였다. semantic segmentation과 이미지 편집, 그들의 "enhancer"네트워크는 생략한다. 따라서 우리는 우리의 생성기를 residual block들로 구성된 다운샘플링 구간과 피쳐맵을 저장하고 이를 입력 해상도로 복원하는 업샘플링 구간으로 완성된 그들의 "global generator"를 사용하여 생성기를 정의한다. 우리는 checkerboard artifacts를 피하기 위해 전치된 컨볼루션을 업샘플-컨볼루션으로 교체하여 약간의 수정을 했다.

High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs loss는 3가지로 구성되어있다. 겐로스, 에프엠로스, vgg로스

모든 생성된 이미지는 입력 이미지에 의존하고, 우리는 이제 의존성을 버리고 표기를 단순화 하

겠다. 전면 및 후면 알베도는 동일한 loss 구성 요소를 사용하지만 전면 및 후면 추정에 대해 별도의 판별자를 사용하여 전문화할 수 있다. 알베도 추정 문제에 이 네트워크를 적용하는 것은 간단하다. 이미지의 합성 훈련 데이터(섹션 4.2 참조)와 해당 전면 및 후면 알베도가 주어지면 두 개의 알베도 세트(그림 4의 중앙)에 해당하는 6개의 채널로 G 를 추정한다. 총 손실은 전면에 적용된 손실의 합계이다.

<3.2 Depth estimation>

적대적 손실로부터 적합하지만 기하를 복구하는 경우에도 마찬가지인 것 같다. 우리의 경험에서 L_{alb} 의 적대적 손실은 깊이 및 정규 추정 문제에 적용될 때 노이즈를 도입하고 보이지 않는 조건에 대한 견고성을 감소시킨다. 이러한 이유로 기하학 추정 목표의 깊이 L_d 및 법선 L_n 항은 적대적 손실을 L_1 손실로 대체한다. L_V G 는 (무제한) 깊이 값의 정규화가 필요하므로 깊이 표현에 적용되지 않습니다. 훈련 불안정을 유발합니다. 총 손실에는 법선 및/또는 깊이에 적용된 이 기하학적 손실과 마스크 출력의 이진 교차 엔트로피 손실이 포함될 수 있다.

<3.3 Estimating Dense Correspondence>

이전 섹션에서 설명한 시스템은 본질적으로 불완전한 픽셀당 깊이 값을 생성합니다. 또한 이러한 값은 픽셀당 생성되기 때문에 의미론적 의미(코, 팔꿈치 등은 어디에 있습니까?)가 없습니다. 이 섹션에서는 인간의 모양과 자세에 대한 통계 모델인 SMPL [29]을 기반으로 신체 기하학의 보이지 않는 부분(그림 3의 검은 부분)을 추론하기 위해 [6]에 설명된 메쉬 정렬 프로세스를 채택합니다.

정렬 프로세스는 자유 몸체 정점 세트(메시라고 함)를 변형하여 이전 섹션에서 추론된 포인트 클라우드(스캔이라고 함)에 가깝고 SMPL 몸체 모델에 따를 가능성도 있습니다. [6]과 유사하게, 우리는 메시로 스캔 거리 항 E_s , 얼굴 랜드마크 항 E_{face} , 두 포즈 및 모양 사전 Epose 및 Eshape, 추론된 자유 정점을 결합하는 항의 가중 평균으로 구성된 손실을 최소화합니다. 모델 E_{cpl} 과 함께. 자세한 내용은 원본 간행물에서 얻을 수 있지만 다음 단락에서 용어에 대한 약간의 직관력을 제공합니다.

E_s 는 스캔과 메쉬 표면의 가장 가까운 점 사이의 제곱 3D 거리에 페널티를 줍니다. E_{face} 는 이미지에서 감지된 얼굴 랜드마크[23](스캔과 암시적으로 일치)와 SMPL에서 미리 정의된 랜드마크 위치 사이의 제곱 3D 거리에 페널티를 줍니다. E_{cpl} 은 자유롭게 변형할 수 있는 메시가 최적화된 포즈 및 모양 매개변수가 암시하는 모델에 가깝게 유지되도록 권장합니다. Epose와 Eshape는 SMPL 매개변수와 CMU 및 SMPL 데이터 세트에서 추론된 가우스 분포 사이의 Mahalanobis 거리에 페널티를 적용하여 결합 모델의 포즈와 모양을 정규화합니다[7].

단일 보기 및 보정되지 않은 다시점 형상 추정에서 일반적이므로 우리의 결과는 대상 축척을 정

확하게 복구할 수 없습니다. SMPL은 임의의 스케일로 스캔을 맞출 수 없으므로 메쉬를 최적화하기 전에 먼저 스캔을 고정 높이로 스케일링한 다음 최적화된 메쉬에 역 스케일을 적용하여 원래 참조 프레임으로 되돌립니다.

깊이 추정기를 훈련할 때 깊이 손실은 전역 제약 조건으로 작용하여 전면 및 후면 스캔이 일관된 규모로 추정되도록 강제합니다. 훈련 중에 이 손실이 생략되면(섹션 4.5 참조) 전면 및 후면 스케일이 반드시 일관성이 있는 것은 아니므로 메쉬 정렬 중에 상대적 스케일을 최적화해야 합니다. 이것은 뒤쪽 정점에 적용되고 메시와 함께 최적화되는 단일 추가 자유 스케일 변수를 도입하여 수행할 수 있습니다. 실험을 설명할 때 이 옵션을 선택 취소라고 합니다.

<4. Experiments>

<4.1 Training and evaluation details>

알베도 추정을 위해 우리는 메모리 제한에 따르기 위해 512X512 random crop을 진행하였다. 다중 스케일 판별기는 1x, 1/2x 및 1/4x 해상도에서 이미지를 처리합니다. 손실은 [46]과 같이 가중됩니다. 깊이 추정을 위해 720 × 960 이미지를 학습하고 초점 길이는 720픽셀입니다. 우리는 카메라까지의 고정된 거리를 가정하지 않습니다. 알베도 및 깊이 추정 네트워크는 모두 배치 크기가 1인 180k 단계로 훈련되고 입력 이미지는 가우스 블러, 가우스 노이즈, 색조, 채도, 밝기 및 대비로 증가됩니다. 훈련 과정은 V100 Tesla GPU를 사용하는 경우 약 48시간이 소요됩니다. 평가는 720 × 960 이미지에서 수행됩니다. 두 네트워크 중 하나의 단일 정방향 패스는 약 100밀리초가 걸리는 반면 SMPL을 스캔에 맞추는 데는 7초가 걸립니다.

<4.2 Datasets>

우리는 합성 데이터 세트(그림 5)에 대해서만 훈련하고 "실험실 내"에서 수집한 실제 이미지에 대해 테스트합니다. 옷을 입고 "A" 자세로 서십시오(그림 7 참조). 40,000개의 합성 이미지 튜플을 렌더링합니다(검증 및 테스트를 위해 각각 1% 보류). 본체에는 SMPL로 합성된 기본 저주파 형상과 실험실에서 캡처한 고주파 변위가 있습니다. SMPL 형상 매개변수는 CAESAR 데이터 세트에서 샘플링되고 포즈는 (a) CAESAR 포즈와 (b) A 포즈에서 편안한 자세까지 팔이 다양한 실험실 내 스캔 포즈 세트에서 샘플링됩니다. 실험실에서 촬영된 사람들의 3D 사진 측량 스캔에서 파생된 텍스처 및 변위 맵은 무작위로 샘플링되어 기본 몸체에 적용되어 입력 및 출력 공간의 다양성을 높입니다.

카메라는 원점에서 0 회전으로 고정되고 몸체는 카메라를 약간 아래쪽으로 기울이면서 약 2미터의 거리를 시뮬레이션하기 위해 무작위로 변환 및 회전됩니다. 구체적으로, 변환은 $x \sim [-0.5, 0.5]$, $y \sim [0.0, 0.4]$, $z \sim [-2.2, -1.5]$ 에서 샘플링되고 회전은 $x \sim [-9.0]$ 에서 오일러 각으로 샘플링됩니다.

35], $y \sim [-7, 7]$, $z \sim [-2, 2]$, yxz 순서로 적용됩니다. 배경 이미지는 사람이 포함된 이미지를 제외하고 OpenImages[25]에서 가져옵니다.

우리는 세 가지 광원을 사용합니다: 이미지 기반 환경 조명(배경 이미지를 광원으로 사용), 점 조명 및 직사각형 영역 조명. 각 렌더에 대해 모든 조명의 강도, 포인트 및 영역 조명의 위치와 색 온도, 영역 조명의 방향과 크기, 바디에 있는 셰이더의 반사도와 거칠기를 무작위로 샘플링합니다. 모든 광원은 레이트레이싱 그림자를 드리우며 일반적으로 영역 및 점 광원에서 가장 잘 보입니다.

<4.3 Visual Evaluation>

기준선으로 L1 손실 함수를 사용하여 정면 깊이의 직접 추정을 고려합니다. 그림 2는 깊이에 대한 L1 손실과 법선에 대한 L1 손실로 훈련된 모델을 비교하여 자연 테스트 이미지에서 추정된 메시지를 보여줍니다. 깊이만 손실한 결과는 사용할 수 없는 것처럼 보이지만 법선만 손실한 결과는 부드럽고 강력하며 인상적인 양의 디테일을 캡처합니다. 따라서 인체의 상세한 깊이 추정의 경우 깊이에 대한 직접적인 손실은 충분하지 않은 반면 표면 법선의 손실은 강력하고 상세한 깊이 추정을 생성하기에 충분합니다. 그러나 법선의 손실은 출력을 로컬로 제한하기 때문에 지오메트리는 스케일에 맞지 않습니다. 깊이의 손실은 기하학의 품질에 중요하지 않지만 그럴듯한 인간 규모의 공간으로 출력을 장려합니다.

팩스의 한 가지 장점은 단일 이미지에서 미묘한 모양 세부 사항을 추출할 수 있다는 것입니다. 복구된 모양은 그림 7의 거의 모든 예에서 허리, 엉덩이 및 가슴에서 관찰된 바와 같이 복잡하고 개인적입니다. 이것은 볼록 껍질[31], 복셀[44] 또는 SMPL 모양 매개변수[22]에 기반한 방법으로는 달성하기 어렵습니다.], [3]과 같이 이미지 윤곽에 맞게 모양을 명시적으로 최적화하는 방법도 기본 최적화가 데이터와 기본(지나치게 부드러운) 모델 사이의 절충안을 찾아야 하기 때문에 이러한 세부 수준을 복구하지 못합니다. FAX에서 얻은 세부 정보는 대부분 윤곽선에서 볼 수 있지만 측면 렌더링은 이 세부 사항이 신체 모양에 걸쳐 일관된 방식으로 재구성되어 실루엣 및 이미지 음영과 일관된 가슴 및 배 모양을 재현함을 보여줍니다.

그림자 및 문신과 같은 시각적 불연속성은 도전 과제입니다. 고전적인 셰이프-프롬 셰이딩(shape-from-shading) 방법은 오해의 소지가 있는 시각적 경계에 능선 인공물을 도입하는 것으로 유명합니다. 그림 7(오른쪽의 3행)에서 볼 수 있듯이 우리의 방법은 문신이 있는 경우 깨끗한 기하학을 생성합니다. 그리고 그림 6에서 우리의 방법은 날카로운 그림자에 대한 불변성을 나타냅니다. 우리는 이 불변성을 훈련 데이터 세트의 다양성에 거의 전적으로 신뢰합니다. 훈련에서 날카로운 그림자를 도입하기 전에(그림 5: 왼쪽의 행 3), 그림자 주변의 능선 아티팩트가 테스트 출력에서 일반적이었습니다.

공간 스캔 구멍은 추가 과제입니다. 많은 고품질 스캐너 설정과 마찬가지로 원시 추정 스캔은 전면 및 후면을 향한 깊이 맵 사이의 이음매로 눈에 띄게 보이는 모든 지오메트리를 캡처하지 않

습니다. 이 문제는 아바타를 맞추는 동기 중 하나입니다. 재배포 가능성을 제공하는 것 이상으로 구멍 폐쇄 및 스캔 완료를 제공합니다. 그림 1과 7은 스캔, 이음새, 구멍 폐쇄를 제공하는 아바타를 보여줍니다.

우리의 전면 알베도 추정 네트워크는 부드러운 그림자에 탄력적입니다. 이를 확인하려면 회색 스캔과 동일한 조명으로 조명되는 그림 7의 RGB 입력 및 전면 텍스처 스캔을 고려하십시오. 특히 오른쪽 4행에서 피부 하이라이트가 제거된 것을 관찰하고 대부분의 다리와 몸통에서 훨씬 더 균일한 피부 톤을 관찰합니다. 7행 오른쪽. 선명하게 드리운 그림자를 제거하는 것은 매우 어렵지만 오른쪽 행 1, 2 및 5에서 합리적인 결과를 얻을 수 있습니다.

우리의 등 알베도 추정기는 피부 톤과 의복 연속성을 포함하여 만족스러운 앞/뒤 일관성을 나타냅니다. 일부 브래지어 끈(예: 7에서 7행 왼쪽)은 연속적이지만 물리적으로 그럴듯하지 않은 구성을 보여 주는 반면 피부색의 의복(7행에서 3행 왼쪽)은 피부 질감과 조화를 이룹니다. 훈련 데이터의 개선은 이 문제를 해결해야 합니다.

<4.4 Quantitative evaluation on Dynamic FAUST>

여러 이미지에서 모양을 추정하는 최첨단 시스템 중 하나인 [3]과 우리 시스템을 정량적으로 비교합니다. [3]에 따라 Dynamic FAUST의 주제에서 합성 렌더링을 생성하고 모양을 추정하고 합성 데이터와 비교하여 평가합니다. [3]과 달리 각 주제에 대해 하나의 이미지만 필요합니다. 또한 우리 시스템은 RGB 이미지로 작동하기 때문에 [8]의 저자는 데이터 세트의 각 주제에 대해 하나의 자연스러운 질감을 친절하게 제공했습니다. [3]에 설명된 절차를 따라 표 1의 오류를 계산합니다. 먼저 섹션 3.2 및 3.3에 설명된 대로 스캔 및 정렬을 추정합니다. SMPL을 사용하여 정렬을 해제하고 크기를 조정하여 실제 모양만큼 높게 만듭니다. 이 고정된 모양을 사용하여 각 메쉬의 정점과 다른 메쉬의 표면 사이의 평균 양방향 거리를 최소화하기 위해 변환 및 크기 조정을 최적화하고 groundtruth에서 변환 및 포즈를 초기화합니다. 더 신뢰할 수 있는 오류 추정치를 얻기 위해 주제당 N개의 합성 이미지에 대해 이 절차를 반복합니다. 이 평균 양방향 거리는 표 1의 왼쪽 열에 보고됩니다. 이 절차는 [3]에 보고된 전체 방법과 비슷합니다. 우리의 오류는 [3]보다 크며, 이는 두 가지 요인에 기인할 수 있습니다. 먼저 [3]이 수백 개의 이미지를 사용하는 동안 단일 이미지에 액세스할 수 있습니다. 둘째, SMPL은 포즈와 모양을 어느 정도 결합하기 때문에 스캔에서 groundtruth 포즈를 적용하는 것은 차선책일 수 있습니다. 이 문제를 분리하기 위해 우리는 또한 표 1의 중간 열에 표시된 스케일 및 변환(항상 고정된 모양 유지)과 함께 포즈를 최적화했습니다. 그러나 이 결과는 [3]과 직접 비교할 수 없다고 생각합니다.

<4.5 Ablation Study>

1. Loss 값들에 대해서 수행

2-1. 네트워크 residual block들의 수 대해서 수행

2-2. downsampling (scales)의 수에 대해서 수행

3. 훈련시 블러처리한 데이터 Augmentation에 대해 수행

- Loss 설정에 있어서 normal, depth 모두 성능에 영향을 미쳤다. 특히, normal(L1, Lvgg)에 대한 중요성이 주목할만 하며, depth Loss를 제거하는 것보다 좋지 않았다.
- residual block 수 줄이면 정확도가 증가하긴 하나, 무시할만한 수준이다.
- 다운샘플링의 수가 적으면 노이즈가 유지되어 정확도가 저하되었다.
- blurr 증대는 수치적 영향은 적지만 스파이크와 구멍을 만들어 텍스처 스캔을 빠르게 생성하는데 사용할 수 없다는 것을 관찰했다.

우리의 방법론의 성능에 기여하는 요소를 연구한다. 먼저 Loss 조건의 개별 기여를 고려한다. 다음으로 네트워크 깊이에 영향을 미치는 네트워크의 잔여 블록 수를 변경한다. 유사하게, 얼마나 많은 다운샘플링 작업(스케일)이 수행되는지 변경합니다. 이러한 작업에는 학습된 컨볼루션이 포함되므로 네트워크에 용량과 깊이가 추가된다. 마지막으로 합성 훈련 데이터에서 수행되는 블러 데이터 증대의 역할을 테스트한다. 우리는 87명의 피험자의 이미지에 대해 이 실험을 실행한다(4개의 피험자 예는 그림 6 참조). 절제 연구의 결과는 표 2에 요약되어 있다. [3]과의 호환성을 위해 섹션 4.4에 설명된 절차를 사용하여 스캔 대신 추정된 정렬을 사용하여 모든 비교를 수행하고 평균 양방향 점 대 메시 거리를 본다. 그러나 스캔에 모델을 맞추면 파이프라인의 덜 견고한 변형(예: "no blurr aug")의 문제가 정규화되고 언포징 프로세스의 불완전성이 미묘하고 잠재적으로 오도하는 부정확성을 유발할 수 있으므로 모델 변형의 절충안은 반드시 이 메트릭으로 잘 표현되어야 한다. opt 포즈로 레이블이 지정된 열은 이전 섹션과 유사하게 거리를 최소화하도록 최적화된 포즈와 관련된다. 또한 깊이가 없는 실험은 이러한 독립적 최적화 없이 정량적 평가를 덜 사용하게 만드는 전면 및 후면의 규모 차이를 보여주기 때문에 전면 및 후면 스케일의 독립적인 최적화(섹션 3.3에 설명된 대로 옵트 백으로 표시됨)를 고려한다. 가장 눈에 띄는 것은 이 손실에서 법선의 중요성이다. 정규화(L1 및 VGG 모두)를 제거하는 것은 깊이 항을 제거하는 것보다 더 해로우며, 이는 그림 2에 제공된 직관과 일치한다. depth 또는 normal을 제거하면 baseline에 비해 부정적인 영향이 발생한다. 다운샘플링을 줄이면 네트워크가 더 얇아지고, 이로 인해 더 자세한 정보와 노이즈를 유지할 수 있어 정확도가 크게 저하된다. blurr 증대는 수치적 영향은 적지만 스파이크와 구멍을 만들어 텍스처 스캔을 빠르게 생성하는데 사용할 수 없다는 것을 관찰했다. 마지막으로 normal에서 VGG 손실을 생략하면 정확도가 약간 손실된다. 그림 6에 추가 구성을 추가한

다. 법선에서 L1 손실을 제거하지만 VGG를 유지하면 더 많은 음영 artifacts가 있는 지나치게 부드러운 스캔이 된다. 마지막으로, residual block의 수를 줄이는 것이 정확도를 향상시킨다는 것은 놀라운 일이지만 우리는 그 차이를 무시할 수 있다고 생각한다.

<5. Conclusions>

FAX는 이전에 볼 수 없었던 세부 수준의 단일 RGB 이미지에서 전신 기하 및 알베도를 추정합니다. 이 품질은 크게 두 가지 주요 요인에 따라 달라진다.

1. 우리는 복셀(voxel), 볼록 껍질(convex hull) 또는 바디 모델과 같은 표현을 통해 출력을 간접적으로 사용하지 않는다. 이를 통해 다른 방법보다 빠르게 이미지 변환 네트워크를 통해 원래 픽셀 정의에서 디테일을 복원 할 수 있다.

2. 우리의 기하 추정(geometry estimation)은 표면 법선(surface normals)의 역할에 크게 의존하며 깊이 정보가 없을 때 표면 법선만으로도 어떻게 그럴듯한 물체를 생성할 수 있는지 보여준다. 두 개의 데이터 세트를 사용하여 시스템을 평가하고 ablation study를 수행하며 시스템의 시각적 성능을 광범위하게 설명한다. 향후 작업을 위해 훈련 데이터를 개선하면 정면 포즈나 최소한의 옷과 같은 현재 방법의 많은 제한 사항을 극복할 수 있다고 믿는다. 우리는 신속하고 데이터 중심적인 방식으로 스캔 기하와 텍스처의 이음새를 제거하고자 한다. 마지막으로, 추가 보기를 통합하면 단일 보기에서 추정된 모양에 존재하는 고유한 모호성을 줄이는 데 도움이 될 수 있다고 믿는다.

추가 공부필요)

- checkerboard artifacts

- High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs

번외) 추가 스크립트

<https://www.amazon.science/latest-news/the-science-behind-the-amazon-halo-band-body-feature>