Journal of the Korea Industrial Information Systems Research Vol.25 No.5, Oct. 2020 :1-11
http://dx.doi.org/10.9723/jksiis.2020.25.5.001 ISSN:1229-3741

의상 이미지의 3차원 의상 복원 방법과 가상착용 응용+

(3D Reconstruction of a Single Clothing Image and Its Application to Image-based Virtual Try-On)

안 희 준[1)*], 미나르 마드울 라흐만[2)]
(Heejune Ahn and Matiur Rahman Minar)

요 약 가상착용기술은 온라인 의류 쇼핑 활성화를 위해 중요한 기술이다. 최근 이미지 기반 가상착용기술은 의상과 착용 대상 신체의 3차원 정보가 필요하지 않다는 실용성 때문에 큰 관심을 받고 있다. 그러나 기존의 이미지 기반 알고리즘의 2차원 기하변형 방식의 한계로 인하여 대상 인물의 포즈와 의상 이미지의 형태가 큰 차이가 있는 경우 자연스러운 의상변형을 하지 못한다. 본 논문에서는 이러한 문제를 해결하기 위해 3차원 인체 모델을 이용하여 2차원 의상 사진으로 부터 의상의 3차원 모델을 생성하고, 대상 인물의 자세와 체형에 맞게 3차원 변형 후 렌더링하고 대상 인간 이미지와 혼합을 통하여 가상착용 이미지를 생성할 수 있다. 기존 연구에서 사용된 VITON 데이터 세트를 사용한 실험 결과는 3차원 변형이 요구되는 경우에 2차원 이미지 기반 가상착용 결과들에 비교했을 때 자연스러운 결과를 보인다.

핵심주제어: 3D 모델 복원, Skinned Multi-Person Linear (SMPL) 모델, 3D 복원, 의상 모델, 가상착용 (VTON)

Abstract Image-based virtual try-on (VTON) is becoming popular for online apparel shopping, mainly because of not requiring 3D information for try-on clothes and target humans. However, existing 2D algorithms, even when utilizing advanced non-rigid deformation algorithms, cannot handle large spatial transformations for complex target human poses. In this study, we propose a 3D clothing reconstruction method using a 3D human body model. The resulting 3D models of try-on clothes can be more easily deformed when applied to rest posed standard human models. Then, the poses and shapes of 3D clothing models can be transferred to the target human models estimated from 2D images. Finally, the deformed clothing models can be rendered and blended with target human representations. Experimental results with the VITON dataset used in the previous works show that the shapes of reconstructed clothing are significantly more natural, compared to the 2D image-based deformation results when human poses and shapes are estimated accurately.

Keywords: 3D Model reconstruction, Skinned Multi-Person Linear (SMPL) model, 3D Reconstruction, Garment model, Virtual Try-On (VTON)

.

* Corresponding Author: heejune@seoultech.ac.kr
+ 이 논문은 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다(2020-0604)
Manuscript received July 14, 2020 / revised August 11, 2020 / accepted August 19, 2020
1) 서울과학기술대학교 전기정보공학과, 제1저자, 교신저자
2) 서울과학기술대학교 전기정보공학과, 제2저자

1. 서 론

온라인 패션 시장은 매년 빠르게 성장하고 있다. 그러나 다른 공산품에 비하여 패션 상품은 착용 대상에 따라 착용 결과나 만족도가 달라지

므로, 착용을 해보기 전에는 구매 결정을 내리기가 쉽지 않다. 따라서 착용 결과를 예측해 볼 수 있는 가상착용 (Virtual try-on: VTON) 기술이 매우 중요하다 (Ahn et al., 2018).

초기의 가상착용기술은 3차원 그래픽스 기법을 그대로 적용하는 형태로 시작되었다. 그러나 3차원 그래픽스 기술을 적용하기 위해 필요한 고객과 의상의 3차원 모델 확보는 일반적으로 쉽지 않고 비용이 많이 든다. 따라서 최근 딥러닝을 사용한 2차원 이미지 기반 기법들이 연구되고 있다. 특히, 의상 사진과 고객 사진을 입력으로 하여 사용자가 해당 의상을 입은 사진을 생성하는 VITON (Han et al., 2018)과 CP-VTON (Wang et al., 2018) 과 같은 연구가 발표되었고, 뒤를 이어 유사 연구들 (Sun et al., 2019; Yu et al., 2019)이 연속적으로 발표되고 있다.

이런 이미지 기반 연구들이 일부 성공적인 결과들을 보이고는 있지만, 아직 여러 한계가 존재한다. 특히, 대상 인물의 자세가 의상의 3차원 변형을 요구하는 경우 매우 어색한 결과를 보인다. 이는 기존의 방식에서 주로 사용되는 TPS (Thin plate spline) (Boolstein et al. 1989)과 같은 2차원 Non-rigid 방식으로는 3차원 변형을 처리하는 것에 근본적인 한계가 존재하기 때문이다. 주의할 점은 기존의 2D 방식들이 논문상으로 상당히 높은 성능을 보이는 것은 데이터셋의 의상이 반소매와 같이 비교적 단순한 형태이기 때문에 3차원 변형의 영향이 크게 눈에 띄지 않기 때문이다.

Fig. 1은 대상인물이 손을 들고 있거나 팔짱을 끼는 등 큰 3차원 변형이 필요한 경우에는 가상착용결과의 화질이 매우 떨어지는 것을 보여준다. 따라서 기존의 방식은 의상의 변형 자체는 실제 목표 의상과는 차이가 매우 크나, 블렌딩 과정을 거치면서 이러한 점이 가려지고 있다.

본 논문에서는 이러한 3차원 의상변형은 2차원 이미지 기법으로는 해결하기 어려운 문제라고 보고, 의상 이미지에서 3차원 모델을 복원하는 방식을 적용하고자 한다. 의상은 사람이 착용하는 것이므로 인체와 유사하며, 따라서 인체 3차원 모델로 부터 형태의 추정이 가능하다. 이렇게 복원된 3차원 의상 모델은 신체 모델과의 연관성을 이용하여 체형과 자세를 변형되었을 때의 의상의 형태를 얻을 수 있다. 따라서 기존의 알고리즘을 통해서 대상의 자세와 체형을 추정할 수 있다면 대상에 가상 착용한 의상을 생성할 수 있다.

본 논문은 다음과 같이 구성하였다. 제 2장에서 본 논문에서 기초가 되는 2차원 가상착용 기술과 3차원 인체모델에 대한 연구, 특히 SMPL (Skinned multi-person linear model) 모델을 바탕으로 한 연구를 소개한다. 제 3장에서는 관련된 2차원 의상 이미지에서 SMPL 3차원 신체 모델을 이용하여 3차원 의상정보를 복원하는 방법을 제안하고, 복원된 3차원 의상모델을 사용자의 자세와 체형에 맞추고 이를 이용하여 변형된 의상을 만들어 사용자 이미지와 통합하는 방법에 대하여 기술한다. 제 4장에서는 구현된 시스템을 기존연구에서 사용한 데이터를 바탕으로 실험한 결과를 비교 제시하고, 제 5장에서 논문의 한계와 향후 연구에 대하여 설명한다.

2. 관련연구

2.1 3차원 인체 모델

본 논문에서는 3차원 신체모델로 SMPL (Loper et al., 2015)을 사용한다. SMPL은 기본 템플릿 메쉬를 바탕으로 하여 23개의 자세와 10개의 체형 변수 벡터 $(\vec{\beta},\vec{\theta})$를 조절하여 다양한 인체를 상당히 정확히 모델링한다. 또한 선형 스키닝 기법을 활용하고 있어서 상용 렌더링엔진으로 구현이 용이하다. 이러한 장점으로 SMPL은 최근에 여러 가지 연구 (Pons-Moll et al., 2017; Zanfir et al., 2018; Weng et al. 2019)에서 활용되었다. 뿐만 아니라 SMPLify (Bogo et al., 2016)와 같이 사람이 포함된 이미지에서 2차원 이미지에서의 조인트와 SMPL의 2차원 투영된 조인트를 매칭하여 SMPL 모델을 추정하는 방법이 존재한다. 여기서 필요한 2차원 조인트는 최근에 개발된 DeepCut

(Pishchulin et al., 2016)이나 OpenPose (Cao et al., 2018)들의 방식으로 상당히 정확한 추정이 가능하다. 또한 최근에는 조인트뿐만 아니라 실루엣을 고려한 방식도 제안되었다.

최근 2-3년 사이에 이러한 신체 모델을 바탕으로 의상을 착용한 사람 전체를 모델링하려는 연구가 있었다 (Bhatnagar et al. 2019). ClothCap (Pons-Moll et al., 2017)은 3차원 스캐너를 사용하여 의상의 3차원 모델을 확보하려는 연구이고, PhotoWake-up (Weng et al., 2019)은 한 장의 사진에서 애니메이션이 가능한 3차원 모델을 형성하려는 연구이다. 또한 본 논문과 독립적으로 올해 의상의 텍스쳐 정보를 복원하려는 연구가 발표되었다 (Mir et al., 2020). 이 연구들과 달리 본 논문은 3차원 모델을 기반으로 의상을 복원하는 것이 아니라 2차원 이미지에서부터 의상의 3차원 정보를 확보하는 방식을 제안한다.

2.2 (이미지 기반) 가상착용 기술

VITON (Han et al., 2018)과 CP-VTON (Wang et al., 2018) 은 이미지를 이용한 대표적인 가상착용 연구들이다. 이 방식들은 의상 이미지와 고객 이미지의 한 쌍을 입력으로 받아 고객의 현재 의상 대신 이미지 속의 의상을 착용시킨 이미지를 생성한다. 이를 위하여 첫 번째 단계로 의상을 사용자의 체형과 자세에 맞도록 변형하고, 둘째로 변형된 의상 이미자와 사용자의 이미지를 합성한다.

의상 변형에 있어서 VITON은 착용 후의 의상의 영역 마스크를 cGAN (Conditional generative adversarial network)를 사용하여 생성하고 이를 SCM (Shape context matching)

Fig. 1 Failures of Image-based VTON Algorithms: Left to Right, Try-on Cloth and Target Human Images, VTON (Warped Clothes and VTON Results), and CP-VTON

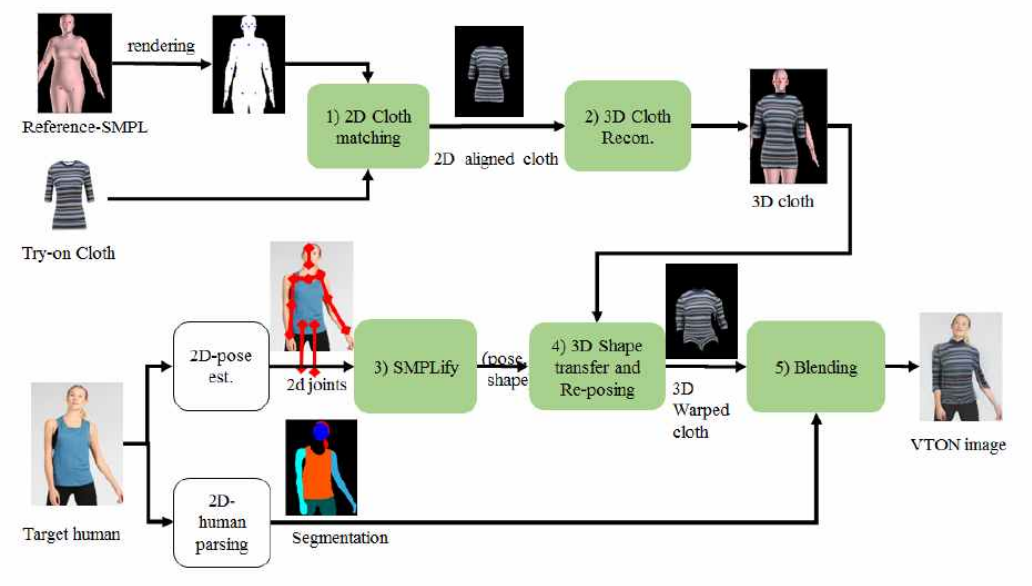

Fig. 2 Proposed Pipeline

(Belongie et al., 2002) 방식으로 매칭한 후 TPS (Thin plate spline) 알고리즘으로 의상을 변형한다. CP-VTON은 CNN Matching 방식에 의하여 TPS 파라미터를 추정하고 이를 이용하는 방식을 사용한다.

그러나 앞서 지적한 바와 같이 현재까지 이미지 기반 연구들은 대상 인물의 자세가 변화가 크거나, 3차원적인 변형이 있는 경우에는 자연스러운 결과를 만들지 못한다.

3. 이미지에서 3차원 의상 모델 복원

Fig. 2는 본 논문에서 제안하는 3차원 의상 모델 복원 방법을 포함한 전체 가상착용 절차를 보여준다. 위쪽 흐름은 새로운 의상이 시스템에 추가되었을 때 동작하며, 아래쪽 흐름은 사용자가 가상착용을 위하여 본인의 사진을 입력하고 입어볼 의상을 선택하였을 때 동작한다.

3차원 의상 모델 복원과 변형 절차는 일상에서 부모가 아이들에게 옷을 입히는 과정에서 착안되었다. 우선 의상을 입히기 전에 옷을 아이의 위에 2차원적으로 대어보고, 아이는 옷을 입히기 좋은 자세로 고정하도록 한다. 이후 아이에게 옷을 입힌 후 아이가 원하는 자세를 취하도록 하여 잘 맞는지를 확인한다. 우선 의상의 3차원 모델을 복원하는 과정은 (1) 2차원상에서 의상과 SMPL 실루엣과 매칭하고 정렬하는 단계, (2) 3차원 인체 모델의 노드들을 2차원상에서 의상영역과 일치시키고, 인체 모델의 깊이 정보를 바탕으로 의상 노드의 깊이정보를 추정/복원하는 단계로 구성된다. 또한 (3) 대상 인물의 자세와 체형을 입력 이미지의 2차원 자세와 실루엣 정보로 부터 SMPLify와 같은 방식으로 추정하고, (4) 인체 모델의 노드를 이에 따라 변형하며 의상모드는 대응 인체모델 노드의 이동 정보를 이용하여 변형한다. (5) 이렇게 얻어진 3차원 의상 모델을 2차원에 렌더링한 후 2차원 이미지 합성방식을 사용하여 결합하여 최종 가상착용 이미지를 생성한다.

3.1 의상 이미지의 2차원 템플릿 매칭

우선, Fig. 3과 Fig. 4의 예와 같이 3차원 SMPL의 파라미터를 고정하고 2차원 렌더링하여 의상 이미지와 2차원 매핑을 할 수 있도록 마스크/실루엣 이미지를 생성한다. 이때 사용하는 SMPL 파라미터는 필요에 따라 변경하여 사용할 수 있으나 본 논문에서는 A-pose를 위한 자세 파라미터와 기본 체형 파라미터를 사용하

였다. 매칭의 정확도를 높이기 위하여 Multi-garment-Net (Bhatnagar et al., 2019)과 유사하게 상의의 종류를 긴팔, 민소매, 3가지 정도의 짧은 팔로 구분하였다.

Fig. 3 Clothing Styles in VITON Dataset: Simple Long and Short Sleeved (Highlighted in a Rectangle: 80% of the Dataset)

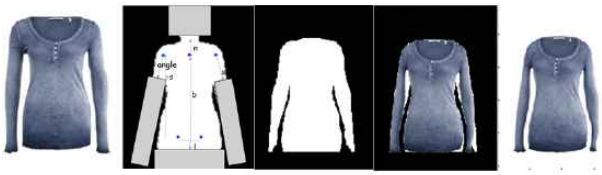

Fig. 4 2D Matching between a Clothing Image and the 3D Body Model Silhouette Mask

Fig. 3에 데이터셋에 포함된 다양한 의상을 보이며, 이 중 본 논문에서의 방법이 적용 가능한 5가지 카테고리로 구분 가능한 90% 이상의 경우에 대해서만 실험하였다. 이후 Fig. 4처럼 각 카테고리에 따라서 인체 실루엣을 자른 후에 의상과 바운더리 점들의 컨텍스트를 기반으로 매칭하는 SCM (Shape context matching) 방식 (Belongie, S et al., 2002)으로 매칭을 수행하고 TPS 변환 $T_{SMPL}(\cdot)$을 통한 변형을 식(1)과 같이 수행한다.

$$(I_{c,warped}, M_{c,warped}) = T_{SMPL(\beta_0,\theta_0)}(I_c, M_c) \quad (1)$$

여기서 I_c과 M_c는 각기 입력 의상 이미지와 의상 영역의 마스크 이미지이고 $I_{c,warped}$는 $M_{c,warped}$는 변형에 의한 출력 이미지이다.

이 단계는 판매자에 의하여 작업되고 실시간성이 필요 없으므로 본 논문에서는 정교한 결과를 위하서 매뉴얼작업을 한다고 가정하였으나, 향후 딥러닝 방식 등에 의한 자동화를 고려할 수 있다.

3.2 3차원 의상 모델 복원

2차원 인체모델에 적절히 맞춰진 의상에서 의상의 3차원 모델을 생성하는 과정은 다음과 같다. 우선, SMPLify (Bogo et al., 2016)에서와 같이 카메라의 프로젝션 변환 $P = K \cdot [R \mid t]$을 이용하여 인체 모델의 노드를 2차원 상에 매핑한다. 여기서 K는 카메라 내부특성 행렬이고, R과 t는 카메라 간 회전, 이동에 대한 외부 파라미터 행렬과 벡터를 나타낸다. 매핑된 노드 중 중심으로부터 멀리 있는 노드들을 외각노드라고 정의하고, 이 노드들과 의상 마스크 M_c의 화소 위치 중에서 가장 가까운 화소를 TPS 변환을 정하는 대응점(매칭점)으로 사용한다. 이렇게 대응된 점들을 사용하여 TPS 변환 파라미터를 추정 후 이를 전체 노드들에 적용 시킨다.

$v_{clothed}$는 3차원상의 의상의 버텍스 v_{body}는 2차원상의 신체의 버텍스라고 표현하면, 이렇게 얻어진 2차원 의상의 좌표값 $T_{TPS}(P \cdot v_{body})$에 대응하는 인체노드가 가지고 있는 깊이 값 $depth(v_{body})$을 합쳐서 의상의 카메라 모델로부터 3차원 좌표값을 식 (2)와 같이 역산할 수 있다.

$$v_{clothed} = P^{-1} \cdot [T_{TPS}(P \cdot v_{body}) | depth(v_{body})] \quad (2)$$

인체의 깊이 값과 의상의 깊이 값은 차이가 있을 수 있기 때문에 구현상으로는 이 간격을 보장하기 위하여 작은 깊이 변화를 추가하였다.

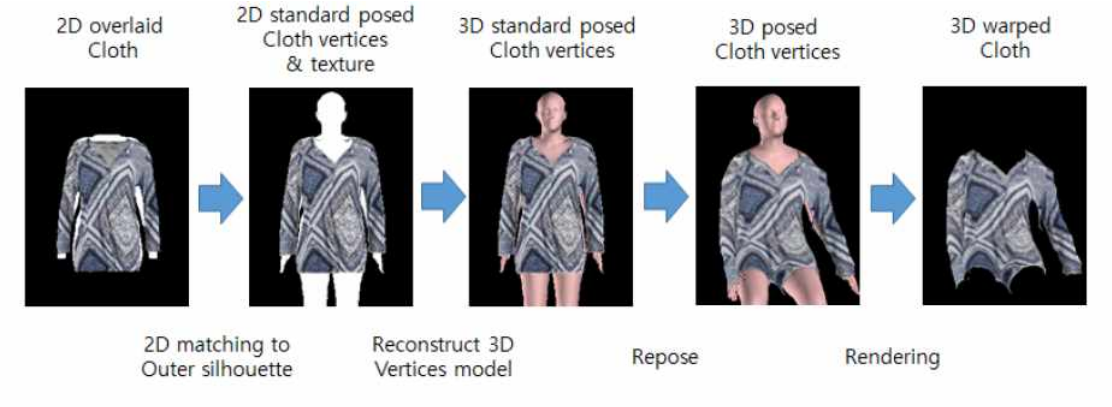

Fig. 5 Flow of Our Method: from 2D Matching of Clothing to 3D Reconstruction and Clothing Transfer

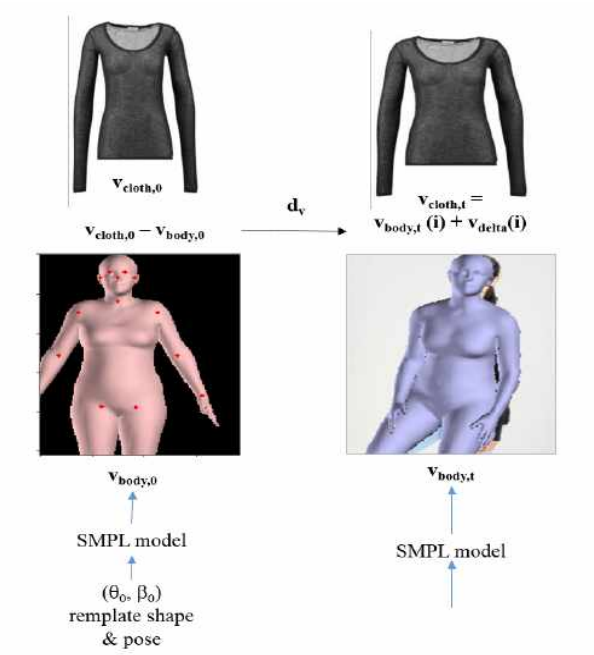

Fig. 6 Clothing Shape and Pose Transfer Method: the Difference between the Corresponding Clothing and Body in the Reference Model are Added to the Target Body Vertex Positions

따라서 현재는 타이트한 형태의 의상의 복원은 높은 성능을 보이지만 헐렁한 옷의 경우에는 개선이 필요하다. 향후 이 부분은 의상의 깊이 예측을 보완하는 방식의 개발을 통하여 향상이 필요한 부분이다. 또한 2차원상의 매칭된 의상 이미지는 메쉬 모델의 텍스쳐 정보로 사용된다.

3.3 사용자 인체 모델 추정

이 단계는 고객의 사진을 바탕으로 SMPL 신체 모델의 체형과 자세 파라미터 $(\vec{\beta}, \vec{\theta})$를 찾아내는 것이다. 본 논문에서 사용한 모델 추정방법은 SMPLify (Bogo et al., 2016)을 따랐으나, SMPLify 방식이 주로 전신사진을 사용한 반면 본 논문에서의 데이터는 반신 사진이므로 사진에서 포함되는 조인트들만을 최적화에 사용하도록 수정하여 사용하였다. 또한 SMPLify 알고리즘의 불완전성으로 부정확하게 매핑되는 결과들이 종종 있는데 (약 20%에서 30% 정도), 이는 사람이 육안으로 검토하여 제거하였으며, 이후 실험에서는 이러한 결과들은 제거하고 실험하였다. SMPLify 알고리즘에 대항 자세한 사항은 원 논문을 참고하기 바란다.

3.4 3차원 의상 모델 변형

앞서 확보한 의상의 3차원 모델은 표준 인체에 대한 것이다. 따라서 해당 고객에 착용하기 위해서는 고객의 체형과 자세에 맞도록 변형이 필요하다. Fig. 5는 본 논문에서 사용한 방식을 적용한 결과를 보여준다. 본 논문에서는 의상의 변형을 위하여 우선 인체 변형을 수행하고 이에 따라 대응되는 의상의 노드의 위치를 이동하는

방식을 사용하였다. 최근 유사한 방식의 연구(Patel et al. 2020)가 발표되었으나, 이 경우는 의상의 통계 특성을 새로 만들어야 하는 어려움이 있는데 이를 무시하고 연구가 발표되었다.

체형과 자세를 전달하는 방식에 있어서, 의상의 기학적 절대적 크기를 그대로 유지하는 방법과 대상 인물의 체형과 자세 따라 변형을 하는가 하는 근본적인 문제가 있다. 물리적인 관점에서는 전자가 맞는 방법이지만 그럴 경우 의상의 크기 별로 처리를 하여야한다는 어려움이 발생한다. 따라서 본 논문에서는 후자 (Fig 6) 로 제한하여 진행하였다.

세부적으로 의상 노드의 이동 양을 계산하기 위해서는 Fig. 7과 같이 '지역좌표'에 의한 변위를 계산 후 이를 다시 변형된 인체 노드위에 적용하는 방식을 사용하였다. 변형 전후 관심 신체 노드 v_{body}를 기준으로 지역 좌표계는 다음과 같이 정의한다. 노드의 평면수직벡터 $\vec{n}(v_{body})$를 z 축으로 정의하고, 이 노드와 연결된 인접 노드 중 가장 작은 인덱스를 갖는 노드의 방향을 x 축으로 정의한다. 그러면 오른손 좌표계에서 나머지 y축 단위 벡터는 외적으로 구해질 수 있다. 이를 정리하면 식(3)과 같다.

$$\begin{aligned} u_z &= \vec{n}(v_{body}) \\ u_x &= u''_x/|u''_x| \\ u_y &= u_z \times u_x \end{aligned} \tag{3}$$

여기서 $u''_x = u'_x - (u'_x \cdot u_z)u_z$ 과 $u'_x = v_{\min(N_v)} - v_{body}$ 이고, 다시 N_v 는 노드 v와 연결된 노드 집합을 의미한다.

이 지역좌표계를 기준으로 변형 전 의상과 신체사이의 변위는 식(4)와 같이 표현된다.

$$d = (dx, dy, dz) = v_{clothed} - v_{body} \tag{4}$$

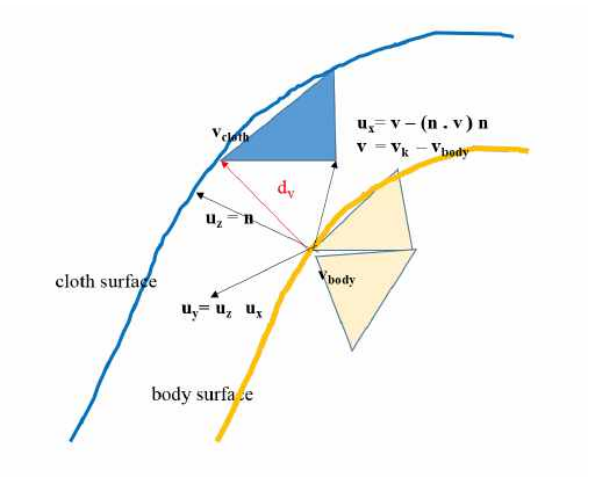

Fig. 7 Local Coordinate Frame Definition for Vertex Displacement Representation.

본 논문에서는 변형된 신체 표면에서의 의상의 변위 좌표는 기존의 신체와 의상간의 변위를 유지한다는 가정한다. 따라서 변형된 의상의 노드의 좌표 $v^t_{clothed}$ 인체 노드에 동일한 변위를 더해서 식 (5)와 같이얻어진다.

$$v^t_{clothed} = v^t_{body} + d \tag{5}$$

실험에 따르면 타이트한 의상의 경우 이 가정이 성공적인 결과를 보였으나, 느슨한 옷의 경우에는 신체의 변형의 정도와 의상의 변형의 정도가 일치하지 않는 문제가 있어서 향후 추가적인 연구가 필요하다.

3.5 변형 의상과 사용자 이미지 합성

3차원 변형된 의상은 렌더링을 수행하여 2차원 이미지와 마스크로 만든 후 고객 이미지와 합성을 수행한다. 이 단계는 CP-VTON의 TOM (Try-on module) 기법을 기반으로 몇 가지 사항을 향상시켜 수행한다. 의상 뿐 아니라 얼굴과 머리 이외의 부분도 입력으로 추가하여 유지하도록 하였고 의상 마스크를 입력으로 넣어서 의상영역의 정보를 명시적으로 제공하였고, 합성 마스크에 대한 손실함수에 식 (6)과 같이 마스크를 포함하였다.

$$L = \lambda_1 \| I - 0 . I_{GT} \|_1 + \lambda_{VGG} L_{VGG} + \lambda_{mask} \| M_{GT} - M_o \|_1 \quad (6)$$

여기서 L_{VGG}는 VGG 손실 (Wang et al., 2018)을, I_{GT}와 M_{GT}는 학습에 사용한 실제착용 이미지와 실제 착용된 이미지에서의 의상 마스크를 의미한다.

4. 실험 및 결과

4.1 구현

실험에는 VITON 연구에서 수집하고 공개한 데이터 셋을 사용한다. 원 데이터는 하나의 의상 이미지와 해당 의상을 입은 모델의 사진으로 구성되어 있으며, 동일한 의상의 조합은 학습데이터로 다른 의상과 조합함으로써 테스트 데이터로 구성된다. 또한 모델 사진은 OpenPose에 의한 2차원 조인트 정보파일과 Look-Into-Person (LIP) (Liang et al., 2018)에 의한 영역 분할 정보 마스크가 함께 제공되며, 의상 사진은 의상 역역의 마스크 이미지가 제공된다. 그러나 원 데이터는 의상과 배경색이 동일하여 마스킹이 잘못 추출된 데이터가 다수 포함되어 있어서 이를 검사하여 총 2032개 중에 1789의 오류가 적은 데이터만을 사용하였다. 또한, 의상과 인물을 통합하는 단계는 CP-VTON에서 사용한 TOM 공개 코드를 일부 수정하여 사용하였다.

4.2 실험 결과

우선 Fig. 8은 3차원 의상 복원 결과의 대표적인 예를 보여준다. 동일한 카메라 시점에서뿐 아니라 다른 시점에서의 결과를 보여준다. 하나의 전면 의상 이미지만을 사용하기 때문에 뒷면은 앞면과 동일한 텍스쳐를 사용하는 것으로 가정하였다. 그러나 실루엣 매칭이 정확하지 않은 경우 측면에서 배경 정보가 포함되는 것을 볼 수 있다. 이러한 문제는 텍스쳐의 경계영역에 의상을 확장함으로써 상당히 감쇠시킬 수 있을 것으로 보인다. 하지만 본 논문에서는 동일한 카메라 위치만 사용하므로 이러한 문제는 최종 결과에 크게 문제가 되지 않는다.

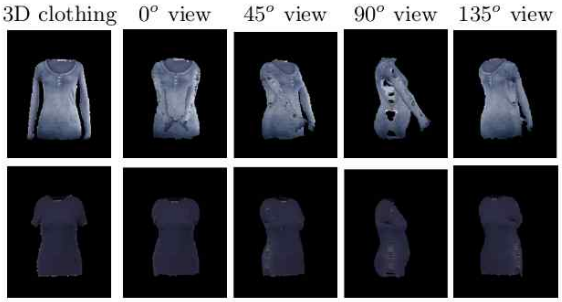

Fig. 8 3D Reconstructed and Transferred Cloth: the First Row for a Long Sleeved Clothing and the Second Row for a Short Sleeved Clothing

Fig. 9에는 대표적인 이미지 기반 방식인 CP-VTON을 사용한 가상착용 결과와 3차원 의상 모델을 통한 결과를 비교 제시하였다. 2018 후반기에 발표된 CP-VTON 이후에 개선한 방식들이 있기는 하지만 기본적인 성능과 특성에는 크게 차이가 없다. 의상의 변형이 작은 경우에는 CP-VTON이 다소 우수한 성능을 보이는 것을 볼 수 있으나, 원 의상 사진의 모습(A-pose 가까움)에서 몸을 크게 비틀거나 팔을 들거나 팔짱을 끼는 등의 의상의 큰 기하 변형이 필요한 경우에서는 제안된 방식이 높은 성능을 보이는 것으로 확인할 수 있다. 동일한 의상을 다시 입히는 경우는 Ground Truth가 현재 대상이미지이므로 정량비교가 가능하다. 본 논문의 주요 제안 사항이 의상의 변형과 최종 VTON 이미지 결과를 비교하려고 하였다. Table 1에서 자세한 비교를 위하여 앞서 5개지의 의상의 종류로 구분하여 IoU (Intersection over union) 을 통해 의상 변형의 성능을 평가한다. 식 (7)은 Ground Truth 영역과 추정된 출력 영역의 겹치는 비율을 계산한다.

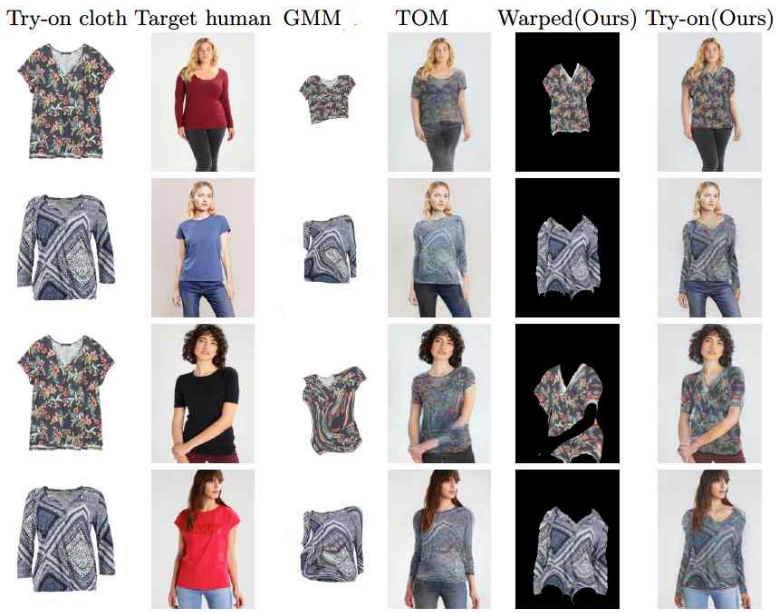

Fig. 9 Qualitative Comparison between the Baseline CP-VTON and Our Approach

Table 1 Quantitative Comparisons between the Baseline CP-VTON and Our Approach

category	Cloth warped		VTON	
	CP-VTON	ours	CP-VTON	ours
long sleeve	0.77	**0.79**	0.76	**0.78**
short elbow	**0.81**	0.78	**0.79**	0.78
short half elbow	0.81	0.81	0.78	0.79
short quarter elbow	**0.82**	0.81	**0.79**	0.78
sleeveless	**0.82**	0.79	**0.79**	0.78

$$IoU = \frac{M_o \cap M_{GT}}{M_o \cup M_{GT}} \tag{7}$$

최종 가상 착용 이미지의 성능은 평균 SSIM (Structural similarity)을 사용한다. 식 (8) 은 두 개의 이미지의 화소들의 편균, 분산, 그리고 구조적인 유사도로 해당 영역의 두 이미지의 유사정을 객관적으로 평가하는 기준 (1이면 동일한 이미지)이다.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{8}$$

그 결과 3차원 변형이 크게 영향을 주는 긴팔 의상의 경우에 제안된 방식이 다소 성능이 앞서는 것을 보인다. 반팔 의상의 경우 이차원 방식에 비하여 다소 성능이 낮은 것은 인체 모델 확보에서 자세측정이 부정확한 경우가 주요한원인인 것으로 보인다.

4.3 제약사항

본 논문의 방식은 처음 시도되는 것으로 향후 개선해야 할 요소들이 많이 존재한다.

첫째, 인체 모델의 깊이 정보만을 사용하여 의상의 깊이 정보를 추정하는 방식은 타이트한 옷의 경우에는 효과적일 수 있는 방법이나, 헐렁하거나 인체와 형태가 다른 부분의 경우에는 바람직한 방식이 될 수 없다. 따라서 좀 더 정확한 깊이 정보를 확보하기 위한 방법이 필요하다.

둘째, SMPLify 또는 유사한 방식의 자세 및 체형 예측 방법은 충분히 정확하지 않으며, 이

런 경우 가상착용 이미지로 통합을 했을 때 인물 영역과 의상영역이 연결이 매끄럽지 못하다. 이 부분은 인체 모델 추정방식의 개선과 함께 추정에 오류가 있을 경우로 통합시 연결이 매끄럽게 하는 방안이 필요하다. 마지막으로 2차원 의상과 표준 인체 실루엣 매칭방식의 성능과 자동화 개선이 필요하다.

5. 결 론

본 논문에서 2차원 의상 이미지에서 3차원 모델을 생성하는 방식을 제안하고 이를 가상 착용 응용에 적용하였다. 실현된 결과를 바탕으로 제안된 방식이 충분히 구현 가능하며 효과적임을 확인하였다. 기존에 대상 인물의 자세가 2차원 방식으로는 변형하기 어려운 경우에는 상대적으로 나은 성능을 보이나, 세부적인 방법의 개선이 필요한 것을 확인하였다.

References

Ahn H., (2018). Image-based Virtual Try-On System: Full Automatic System Design and its Performance, *Journal of Korean Computer Game Society*, 31(3), 37-45.

Belongie, S., Malik, J., and Puzicha, J. (2002). Shape Matching and Object Recognition using Shape Contexts, *IEEE Trans. on PAMI* 24(4), 509-522, https://doi.org/10.1109/34.993558

Bhatnagar, B. L., Tiwari, G., Theobalt, C., and Pons-Moll, G. (2019). Multi-garment Net: Learning to Dress 3d People from Images, *Proc. of the IEEE International Conference on Computer Vision*. pp. 5420-5430.

Bogo, F., Kanazawa, A., Lassner, C., Gehler, P., Romero, J., and Black, M. J. (2016). Keep It SMPL: Automatic Estimation of 3d Human Pose and Shape from a Single Omage, *European Conference on Computer Vision*, pp. 561-578

Bookstein, F. L. (1989). Principal Warps: Thin-plate Splines and the Decomposition of Deformations, *IEEE Trans. on PAMI* 11(6), 567-585, DOI: 10.1109/34.24792

Cao, Z., Martinez, G. H., Simon, T., Wei, S. E., and Sheikh, Y. (2018). Realtime Multi-person 2d Pose Estimation using Part Affinity Fields, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 7291-7299.

Han, X., Wu, Z., Wu, Z., Yu, R., and Davis, L. S. (2018). VITON: An Image-based Virtual Try-on Network, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition,* pp. 7543-7552.

Liang, X., Gong, K., Shen, X., and Lin, L. (2018). Look into Person: Joint Body Parsing and Pose Estimation Network and a New Benchmark, *IEEE Trans on PAMI.* 41, 871-885. https://doi.org/10.1109/TPAMI.2018.2820063

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., and Black, M. J. (2015). Smpl: A Skinned Multi-person Linear Model, *ACM Trans on Graphics,* 34, 248-248. https://doi.org/10.1145/2816795.2818013

Mir, A., Alldieck, T., and Pons-Moll, G. (2020). Learning to Transfer Texture from Clothing Images to 3d Humans, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7023-7034.

Patel, C., Liao, Z., and Pons-Moll, G. (2020). Tailornet: Predicting Clothing in 3D as a Function of Human Pose, Shape and Garment Style, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7365-7375.

Pishchulin, L., Insafutdinov, E., Tang, S., Andres, B., Andriluka, M., Gehler, P. V., and Schiele, B. (2016). Deepcut Joint Subset

Partition and Labeling for Multi Person pose Estimation, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4929-4937

Pons-Moll, G., Pujades, S., Hu, S., and Black, M. J. (2017). Clothcap: Seamless 4d Clothing Capture and Retargeting, *ACM Trans. on Graphics.* 36(4), 1-15 https://doi.org/10.1145/3072959.3073711

Sun, F.D., Guo, J., Su, Z., and Gao, C. Y. (2019). Image-based Virtual Try-on Network with Structural Coherence, *Proc. of the IEEE International Conference on Image Processing*, pp. 519-523

Wang, B., Zhang, H., Liang, X., Chen, Y., Lin, L., and Yang, M. (2018). Toward Characteristic Preserving Image-based Virtual Try-on Network, *European Conference on Computer Vision*, pp. 589-604.

Weng, C. Y., Curless, B., and Kemelmacher-Shlizerman, I. (2019). Photo Wake-up: 3d Character Animation from a Single Photo, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5908-5917

Yu, R., Wang, X., and Xie, X. (2019). VTNFP: An Image-based Virtual Try-on Network with Body and Clothing Feature Preservation, *Proc. of the IEEE International Conference on Computer Vision*, pp. 10511-10520.

Zanfir, M., Popa, A. I., Zanfir, A., and Sminchisescu, C. (2018). Human Appearance Transfer, *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 5391-599.

안 희 준 (Heejune Ahn)

- 종신회원
- KAIST 전기정보공학과 박사 (2000)
- (주) LG전자 차세대단말연구소 선임연구원(1998-2002)
- (주) Tmax 소프트 책임연구원 (2002-2004)
- 서울과학기술대학교 전기정보공학과 (2004-현재) 정교수
- 관심분야 : 컴퓨터비젼, 컴퓨터 통신, 데이터 마이닝

미나르 마드올 라흐만 (Matiur Rahman Minar)

- BUET (방글라데시) 컴퓨터공학과 학사 (2015)
- Automation Solutionz Inc., (캐나다) 2014-2018, 방글라데시 원격근무 프로그래머
- 서울과학기술대학교 전기정보공학과 (2019-현재) 석사과정
- 관심분야 : 컴퓨터비젼, 딥러닝, 데이터마이닝