

# 이미지에서 3차원 인물복원 기법을 사용한 패션쇼 애니메이션 생성기법<sup>+</sup>

(Fashion-show Animation Generation using a Single Image to 3D Human Reconstruction Technique)

안 희 준<sup>1)</sup>\*, 미나르 마드올 라흐만<sup>2)</sup>  
(Heejune Ahn and Matiur Rahman Minar)

**요 약** 본 논문은 단일 이미지를 패션쇼 워킹 영상으로 변환하는 기술을 소개한다. 일반인이 가상으로 패션모델이 되어 보는 흥미로운 응용일 뿐 아니라, 나아가 가상 착용기술과 함께 결합하게 되면 의상착용결과의 동적인 확인이 가능한 기술이다. 본 논문에서 사용한 기술은 이미지에서 3차원 인간신체 모델을 추정 복원해 주는 SMPLify 기법에 기초하여, 인체 모델에서 의상을 포함한 사람으로 모델을 확장하고, 이에 애니메이션 기법을 적용하여 구현되었다. 인체와 의상을 포함한 사람의 3차원 모델은 2차원 이미지 상에서 기하변형과 깊이정보를 사용하여 복원하였다. 패션 데이터 셋에 적용해 본 결과 정자세의 경우에는 성공적인 수준의 결과를 보였으나, 상용수준의 성능을 위해서는 이미지의 분할 기술, 매핑기술 및 가려진 영역의 복원기술 등 선 후처리 기술에 보완이 필요한 것으로 확인되었다.

**핵심주제어:** 인버스-모델, 컴퓨터 그래픽스, 애니메이션, 패션쇼, SMPLify

**Abstract** In this paper, we introduce the technology to convert a single human image into a fashion show animation video clip. The technology can help the customers confirm the dynamic fitting result when combined with the virtual try on technique as well as the interesting experience to a normal person of being a fashion model. We developed an extended technique of full human 2D to 3D inverse modeling based on SMPLify human body inverse modeling technique, and a rigged model animation method. The 3D shape deformation of the full human from the body model was performed by 2 part deformation in the image domain and reconstruction using the estimated depth information. The quality of resultant animation videos are made to be publically available for evaluation. We consider it is a promising approach for commercial application when supplemented with the post - processing technology such as image segmentation technique, mapping technique and restoration technique of obscured area.

**Keywords:** Inverse modeling, Computer graphics, Animation, Virtual fashion show, SMPLify

\* Corresponding Author: heejune@seoultech.ac.kr

+ 이 논문은 본 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다(2018-0531)

Manuscript received August 06, 2019 / revised August 31, 2019 / accepted September 09, 2019

1) 서울과학기술대학교 전기정보공학과, 제1저자, 교신저자

2) 서울과학기술대학교 전기정보공학과, 제2저자

## 1. 서론

논문 (Ahn, 2018a; Ahn, 2018b)에 따르면 온라인 패션 시장 규모는 매년 급속히 증가하여, 미국의 경우 2017년 32조 규모이며, 유럽의 경우 16조

원규모로 추산된다. 국내의 경우 2017년 추산 약 6000억 원으로 전체 온라인시장의 10%에 달하고, 매년 5~10%정도의 증가율을 보인다. 전자제품 등과 달리 의류는 규격화가 쉽지 않고 형태와 색상, 재질이 매우 다양하며, 개인적인 선호도의 차이가 크기 때문에 온라인에서 사진만으로 구매결정이 어렵고 저가 상품에 제한되어 있다. 따라서 온라인 가상 착용기술이 판매품목의 고가화와 양적 성장을 위하여 반드시 필요하다. 이러한 필요성으로 인하여 최근 가상 착용기술이 연구 개발되었으며, 실제적으로 의상과 사람 신체의 3차원 모델을 확보하기 어렵기 때문에 2차원 이미지를 사용한 가상 착용기술이 연구되고 있다(Ahn, 2018a; Ahn, 2018b). Fig. 1에 대표적인 가상 착용 기술에 대하여 예시를 보였다.

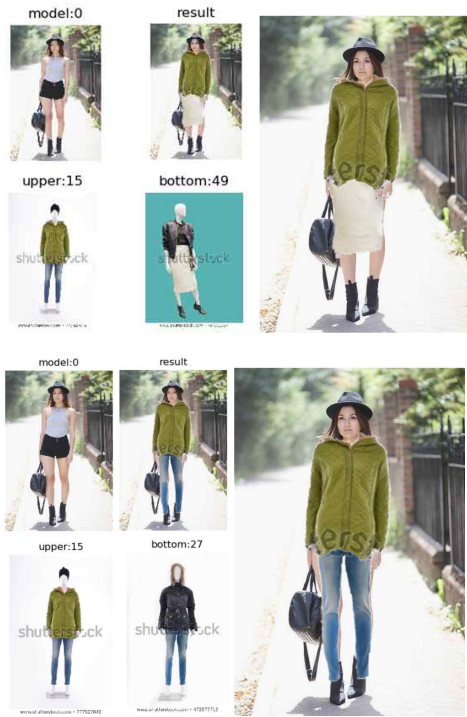


Fig. 1 The Input and Final VTON Image (enlarged) (Left: Upper & Skirt, Right: Upper & Pants) (Ahn, 2018b)

본 논문은 한 장의 의상을 입은 전신 이미지로 패션쇼에 모델로 동작하는 애니메이션 영상을 만드는 기술을 제안한다. 논문은 앞서 계속 연구 중

인 가상착용기술에 대한 연구의 과정에서 사용자가 의상을 입었을 때 단순히 하나의 이미지 이상의 다양한 각도와 동작에서 어떻게 보여 질 것인가 하는 것을 평가하기 위한 목적으로 시작되었다. 현재는 그 중간 결과로 의상 교체 없이 이미지의 대상으로 부터 3차원 모델을 복원하고 이를 애니메이션 하는 기술을 개발하였다. 본 논문에서는 애니메이션 예로 패션쇼만을 대상으로 하지만, 다른 형태의 애니메이션으로의 변경도 기본 알고리즘의 변화 없이 가능하다.

3차원 인체모델과 생성은 그 응용의 중요성으로 매우 방대하게 연구되어 왔다. 본 논문에서 전부 언급하기는 어려우므로, 본 논문의 기초인 SMPLify(Bogo et al., 2016)에 설명된 관련 연구를 참조하길 바란다. 본 연구에서 중요한 기술적인 요소는 인체모델과 이미지를 바탕으로 의상까지 포함하는 사람의 3차원 메시를 복원하는 부분이다. 본 연구와 가장 유사한 연구는 2018년 발표된 Human Appearance Transfer (Zanfir et al., 2018)과 2019년에 발표된 PhotoWakeup (Weng et al., 2019)논문이다. 두 논문 모두 SMPLify에 기반하고 있고, 하나의 이미지를 바탕으로 3차원 사람 모델을 추출해 내고 있다. 그러나 세부적인 변형 알고리즘이나 3차원 구성방식에서 차이가 있다. 방법론적으로 보아 PhotoWakeup은 신체부분을 하나씩 모델링하여 합치는 방식을 적용한 반면 본 논문에서는 매핑방식을 달리하여 머리 부분과 몸통으로만 구분하여 처리하였고, TPS 변환을 사용함으로써 비선형적 변형에 강인하도록 하였다. 또한 기본 포즈로 만드는 알고리즘도 추가로 제시되었다. 아쉽게도 두 논문 모두 데이터와 코드를 공개하고 않고, 적용 데이터 셋에 차이가 있어서 성능상의 비교를 하기는 어렵다.

본 논문의 구성은 다음과 같다. 제2절에서 제안하는 시스템의 전체 구조와 기능에 대하여 설명한다. 우선 전체 시스템의 구성에 대하여 설명하고, 2차원 조인트로부터 3차원 인체모델을 복원하는 SMPLify (SMPL: Skinned Multi-Person Linear Model) 기술에 대하여 설명한 다음 SMPL모델 (Loper et al., 2016)를 기반으로 의상을 포함한 사람 전체의 3차원 모델을 복원하는 기법에 대해

여 설명하고, 마지막으로 복원된 사람모델을 사용한 애니메이션 기법에 대하여 설명한다. 제3절에서 샘플 이미지를 통하여 실험한 단계별 결과 및 최종 애니메이션 영상을 보인다. 결과 제시는 정성적인 성능평가를 같이하며, 특히 제안된 방식의 문제점들을 분석한다. 결론에서 발견된 문제점들에 대한 장-단기적 해법에 대하여 논거하고, 나아가 본 논문이 가상 착용기술과 결합하였을 때 어떤 결과물을 나올 수 있을지에 대하여 기술한다.

## 2. 제안 방법

### 2.1 전체 시스템 구성

Fig. 2는 제안하는 시스템의 전체 구성을 보여준다. 첫 단계에서는 2차원 이미지에서 사람 영역을 분할하고 사람의 골격정보인 조인트 정보를 추정한다. 다음 단계는 대상의 인체 3차원 모델을 SMPLify 방식으로 복원하는 단계이다. 이렇게 획득한 인체 모델과 2차원 분할된 의상을 포함한 사람의 마스크를 사용하여 전체 사람의 3차원 메시 모델을 획득하고, 최종적으로 이 3차원 모델을 골격을 사용하여 Kinematics 기법을 적용하여 원하는 애니메이션 동영상을 만들게 된다.

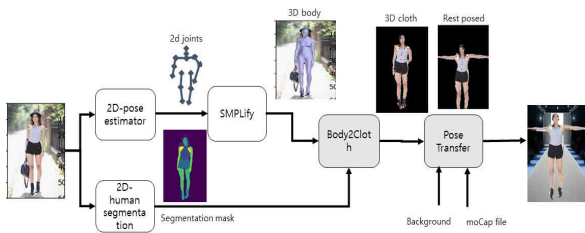


Fig. 2 The Image to Fashion Animation System

### 2.2 SMPLfy기반 인체 이미지 기반 3차원 모델 복원 기법

인체의 모델링 기법은 의료, 행동분석, 스포츠공학, 영상 보안 등 다양한 분야에 응용성이 높아서 지속적인 관심과 발전을 이루어 왔다. 연

구의 분야도 다양하게 자세 예측, 의상, 동작 예측 등으로 연구되어왔다. SMPLify (Bogo et al., 2016)기법은 독일 막스플랑크 연구소에서 2016년에 발표한 SMPL 3차원 인체 모델의 3-D 골격위치의 프로젝션과 2-D 골격 좌표를 매칭하여 SMPL모델의 자세 변수와 형태 변수를 추정하는 방식을 사용한다.

SMPL 모델은 인체의 3차원 모델을 자세  $\vec{\theta}$ 와 체형  $\vec{\beta}$ 에 대한 파라미터만 가지고 일반화한 모델로, 애니메이션이 가능하도록 상용화된 Skinning 알고리즘에 적용이 가능하도록 개발되었다. 기존의 인체모델은 보통 특정인물을 대상으로 모델링을 한 반면, SMPL은 파라메타를 조절하여 새로운 대상에 적용할 수 있는 점이 장점이다. Fig. 3을 SMPL 모델 구성의 단계를 설명한다. 우선 식 (1)과 같이 기본 3차원 메시  $\bar{T}$ 에 자세와 체형에 의한 변화를 반영( $B_s(\vec{\beta}) + B_p(\vec{\theta})$ )하는 첫 단계와 식 (2)의 조인트  $J(\vec{\beta})$ 를 고려한 자세변화에 의한 Skinning  $W(\cdot)$ 으로 구성된다. 보다 자세한 SMPL에 대한 설명은 Loper et al. (2016)을 참고하기 바란다.

$$T_p(\vec{\beta}, \vec{\theta}) = \bar{T} + B_s(\vec{\beta}) + B_p(\vec{\theta}) \quad (1)$$

$$M(\vec{\beta}, \vec{\theta}) = W(T_p(\vec{\beta}, \vec{\theta}), J(\vec{\beta}), \vec{\theta}, W) \quad (2)$$

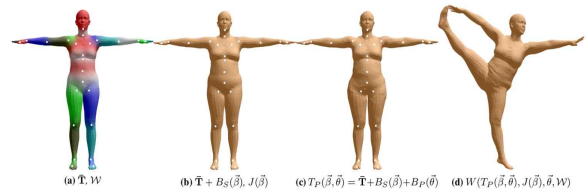


Fig. 3 SMPL Model Parameters and Formation (Loper et al., 2016)

SMPLify는 SMPL 모델이 자세  $\vec{\theta}$ 와 체형  $\vec{\beta}$ 의 변수벡터의 함수로 표현된다는 점을 이용하여 2차원 이미지에서 사람의 조인트 값  $J_{est}$ 과 가장 잘 일치하는  $J(\vec{\beta}, \vec{\theta}, \bar{T})$ 가 되도록 비용함수를 최소화하는 변수를 찾는다. 이때 최적화 비용함수는 식 (3)과 같이 정의된다.

$$E_{tot} = E_J(\vec{\beta}, \vec{\theta}; K, J_{est}) + \lambda_\theta E_\theta(\vec{\theta}) + \lambda_a E_a(\vec{\theta}) + \lambda_{sp} E_{sp}(\theta; \vec{\beta}) + \lambda_\beta E_\beta(\vec{\beta}) \quad (3)$$

식 (3)의 첫 번째 항  $E_J(\vec{\beta}, \vec{\theta}; K, J_{est})$ 은 카메라 변수  $K$ 로 렌더링한 SMPL 모델의 조인트위치와 주어진 2차원 조인트위치와의 차이이고, 나머지 항들은 사람의 자세나 체형에 대한 전제 비용 (Prior)들인데, 두 번째 항  $\lambda_\theta E_\theta(\vec{\theta})$ 은 사람의 조인트의 방향성, 세 번째항  $\lambda_a E_a(\vec{\theta})$ 는 통계적인 자세, 마지막 항  $\lambda_\beta E_\beta(\vec{\beta})$ 은 통계적인 체형의 비중을 반영한다. 네 번째 항  $\lambda_{sp} E_{sp}(\theta; \vec{\beta})$ 은 사람의 신체가 서로 교차할 수 없음 (Spatial penetration)을 반영하는 항이다. 보다 자세한 설명과 예시 및 일반적인 데이터셋에 대한 성능에 대한 분석은 Bogo et al. (2016)을 참고하기를 바란다.

SMPLify 계산을 위해서는 우선 2차원 이미지에서 골격 정보를 얻어야 하는데, 사람이 직접 입력할 수도 있지만 최근에 딥러닝 기법을 사용하여 이미지로부터 골격을 예측하는 기술이 상당한 정확도 (95% 이상)를 보이기 때문에 이를 사용한 자동화된 결과를 활용하였다. 이 논문에서는 OpenPose (Cao et al., 2017)을 사용하여 추출한 골격을 사용하였다. 근본적인 측면에서 3차원 골격과 2차원골격의 일치성 추정된 2차원 골격의 정확성 등에 문제가 있을 수도 있으나, 이 부분은 본 논문의 대상에서 벗어나는 것으로 충분한 정확성이 만족된 것으로 보고 여기서는 더 이상 고려하지 않는다.

### 2.3 SMPL 모델에서 의상을 포함한 사람전체 모델 복원

전 단계에서 얻어진 SMPL모델은 의상을 고려하지 않은 인체모델이다. 의상은 형태가 매우 다양하고 예측하는 것이 현재로는 어렵다. 이 논문에서는 입력 이미지의 사람의 실루엣을 사용하여 의상의 형태를 복원하는 방법을 사용한다. 우선 사람의 실루엣을 얻는 방법은 논문 (Ahn, 2018a)에서와 같이 딥러닝 사람분할기법을 사용하여 구하는데, 본 논문에서는 DeepLab2 (Chen et al., 2017)를 이용하였다.

본 논문에서는 주어진 2차원 이미지의 사람의

실루엣 정보를 이용하는 방법을 개발하였다. 단계 1) 우선 이를 위하여 인체 모델의 3D 메시 노드를 카메라 모델을 사용하여 2차원 이미지에 투영한다. ( $M_{2D}^B = K \cdot M_{3D}^B$ ). 단계 2) 이렇게 얻어진 2차원 노드들을 이미지의 마스크 전체를 커버하도록 2차원 상에서 변형한다. 본 논문에서는 경계 노드들과 마스크의 실루엣 그리고 앞서 구한 조인트를 사용하고 이차원 변형방식인 TPS (Thin Plate Spine) (Bookstein et al., 1989) 변환을 사용하였다. 우선 2차원 투영된 경계점들이 2차원 이미지의 실루엣과 매칭 되도록 TPS 변수를 추정한다. 즉,

$$Warp^*_{TPS} = \min |Warp_{TPS}(BM_{2D}^B) - IM| \quad (4)$$

여기서  $BM_{2D}^B$ 는 이차원으로 인체 모델이 투영되었을 때의 경계노드들이고,  $IM$ 는 대응되는 이미지의 실루엣 경계점들이다. TPS 경계점은 실루엣의 점들 (일정한 간격으로)에서 가장 가까운 노드를 찾는 방식으로 사용하였다. 단계 3) 이렇게 얻어진 TPS를 사용하여 2차원 노드를 모드 변형한다. ( $M_{2D}^H \leftarrow Warp_{TPS}(M_{2D}^B)$ ). 단계 4) 3차원 신체 메시 노드와 카메라의 위치를 사용하여 깊이 정보를  $D_i$ 을 계산한 후 이를 변형된  $M_{2D}^H$  노드의 이미지좌표와 합쳐서 3차원 좌표를 확보한다.  $M_{3D}^H = K^{-1}(M_{2D}^H D)$ . 단계 5) 최종적인 렌더링은 OpenGL 방식의 텍스처 렌더러를 사용한다. 3차원 모델의 색상정보는 이미지를 사용하고, 프로젝션 정보를 이용하여 텍스처의 메시 모델로 사용한다.

이 부분이 본 논문에서 새롭게 추가된 내용이므로 설계에 고려한 사항에 대하여 좀 더 설명하기로 한다. 단계 4의 깊이정보를 추정하는 방식은 위에서 정의한 방식 이외에 변형된 만큼 비례하여 깊이를 정하는 방법을 사용해 보았으나 품질상의 의미 있는 차이를 보이지는 않아서, 계산이 추가로 필요하지 않도록 인체의 깊이정보를 그대로 사용하는 방식을 사용하였다. 의상이 타이트하지 않은 경우에는 인체와 의상 간에 깊이 차이가 크다. 이는 향후 개선해야 할 부분으로 보인다. 최근에 딥러닝을 사용한 단일 영상을 사용한 깊이추정기법 (Liu et al., 2015)들이 개발되고 있으나, 주로 자율

주행 등 응용을 대상으로 하고 있으며, 의상에 적용할 만큼의 정밀도를 확보하고 있지 못한 것으로 판단하여 적용하지 않았다.

위의 단계에서 단계 2가 품질에 있어서 매우 중요한데, 우선 대응점을 정할 때 노드에서 마스크 실루엣을 찾는 방법은 마스크에 노드에 필요한 모든 매칭점이 없을 수 있기 때문에 불가능한 방법이다. 유사한 논문인 PhotoWakeup에서는 영역별로 노드와 경계를 구별하였으나, 이는 자동화가 현재로는 어려운 방법인 것으로 판단된다. 그러나 매칭점이 크게 차이나는 경우 해당 노드 이외에 노드가 영향을 받음을 보였다. 이를 해결하기 위하여 머리 부분과 몸통부분을 나누고 각기 따로 TPS 모델과 적용을 하여 이를 합치는 방법을 사용하였다. 또한 PhotoWakeup은 2차원 변형알고리즘도 Mean Value Coordinate 방식을 사용하였다.

### 2.4 자세 변형 및 애니메이션

SMPL 모델 자세와 체형 벡터를 결정하면 템플릿을 기반으로 하여 모든 메시들의 좌표 값이 결정되어진다. 앞서와 같이 의상을 포함한 사람의 모델은 더 이상 자세와 체형 벡터로 정의되지 않지만, 여전히 조인트를 기준으로 한 Skinning 기법 (Parent, 2012)을 적용할 수는 있다. 동작파일은 최근 애니메이션 업체나 사람의 동작을 연구하는 기관들에 의하여 MoCap이라는 이름으로 이러한 파일을 제공하고 있다. 이 논문에서는 이를 이용하여 동작파일의 조인트 정보를 적용하여 패션쇼에서 걷기 하는 애니메이션 영상을 제작한다.

BVH 파일은 각 조인트들의 관계가 계층적으로 정의되어 있고, 각 조인트들의 기본 각에서 변위값 들을 일정시간 간격으로 저장하고 있다. 입력 이미지상의 사람의 자세는 초기 기본 조인트 각과 다른 값을 가지고 있기 때문에 BVH파일을 적용하기 위해서는 기본자세로 일단 변형을 해줘야 한다. 기본자세로 만들기 위해서는 현 자세의 각도들을 역으로 적용하여야 하는데, 간단하지만 역방향 Kinematics를 풀어야 한다. 이때 주의할 것은 조인트의 각을 적용 시, 전방향 Kinematics를 적용할 때는 말단 조인트부터 변형을 적용하지만, 복원을 할 때는 역과정을 거쳐야 함으로 루트 조인트부터 적용하여야 정상적인 결과를 얻을

수 있다. Fig. 4에 적용원리를 그림으로 표현하였다. 또한 Fig. 5에 적용 예를 보여주고 있는데, 첫 번째 예는 역과정에서 조인트 변환 과정을 전방향 Kinematics과정과 같게 적용한 경우이고, 두 번째 그림이 역순으로 바르게 적용한 결과이다.

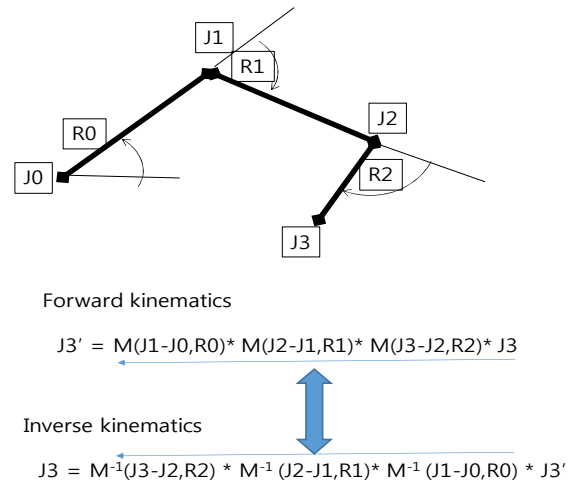
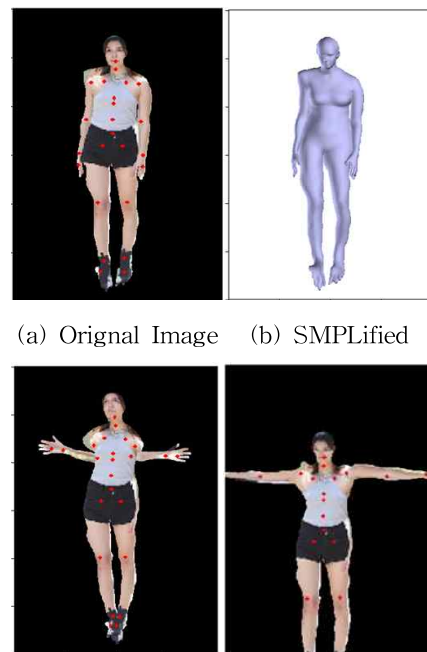


Fig. 4 Rest Pose Construction Formula



(c) Wrong Inverse (d) Correct Inverse  
Fig. 5 Rest Pose Construction Examples



이렇게 얻어진 기본포즈로 모델을 기준으로 하여 LBS 방식으로 애니메이션을 수행하였다. 패션쇼와 유사한 환경을 위하여 배경은 패션쇼 무대 사진을 이용하였고, 카네기 멜론대학에서 생성한 걷기자세를 BVH (Bio Vision Hierarchy)(Guo et al., 2014)파일로 변형한 내용을 사용한다. BVH 파일의 움직임의 방향이 정면이 아니기 때문에 Offset 은 사용하지 않았으며, 대신 카메라의 위치를 조절하여 이동을 시뮬레이션 하였다. 동영상 결과의 샘플화면은 Fig. 8에 제시한 바와 같으며, 실제 동영상은 해당 URL에서 확인할 수 있다.

### 3. 실험 결과

본 논문에서 사용한 이미지 데이터들은 (Ahn, 2018a)에서 사용한 패션 이미지 데이터 셋을 사용하였다. 이 데이터 셋은 10,000개의 주로 여성의 패션모델사진으로 구성되어 있다. 모델의 자세는 앞을 보고 서있는 자세부터, 옆으로 서있는 자세, 앉아 있는 자세 등 매우 다양하다. 본 논문에서는, 전면 의상확보가 가능한 정면으로 서있는 자세를 중심으로만 실험하였다. 하지만 정면자세에도 팔이 몸을 가르고 있거나 다리를 교차하고 있는 자세 등이 포함되어 있고, 모자를 쓰고 있거나 긴 머리로 몸의 일부를 가리는 등 쉽지 않는 문제요소를 가지고 있어서, 이러한 요소가 결과에 어떤 영향을 미치는지, 현재 까지 개발된 기법들로 어느 수준까지 해결이 가능한지를 확인하고자 하였다.

#### 3.1 SMPL 모델 적용 결과

SMPLify 논문에서는 알고리즘이 다양한 자세에서 적용가능한지를 확인하기 위하여 스포츠 활동 LSP (Leeds Sports Pose) 데이터 셋을 사용하였으나, 본 논문에서는 상대적으로 자세가 단순한 패션모델들에 의한 이미지를 사용하였다. 정자세에 가까운 이미지의 경우는 상당히 우수한 자세매칭이 됨을 확인하였다. 하지만, 특히하게도 전반적으로 다리부분이 짧게 예측되는

경우가 많이 발생하였다. 이러한 현상은 두 가지 원인으로 생각해볼 수 있는데 하나는 일반인들을 상대로 한 통계분포에 비하여 실험에 사용한 패션모델들의 경우 다리가 긴 체형인 경우가 많기 때문인 것과 SMPLify 최적화에서 2차원 골격 추정의 신뢰도를 최적화 알고리즘에 사용하는데, 2차원 추정에서 신뢰도가 떨어지기 때문인 경우로 볼 수 있다. 실제 두 가지 원인을 확인하기 위해 데이터베이스를 통하여 데이터를 확인하였다. Fig. 6에 대표적인 경우 각각 두 개씩의 데이터를 비교하였다. a, b는 다리 조인트 매칭이 잘 된 경우이고 c는 약간, d는 심하게 짧게 매칭된 경우이다. 분석결과 1번의 가능성이 큰 것으로 확인되었다.

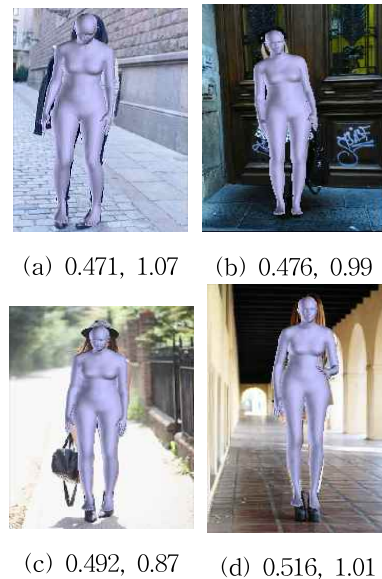


Fig. 6 SMPLify Result Analysis (Fashion Dataset) (Leg/Height, Ankle/Confidence)

#### 3.2 의상을 포함한 사람 전체의 복원 결과

우선 2차원에 투영된 SMPL 모델은 2차원 상으로 투영되면서 단일 표면이 아니라 여러 겹으로 겹쳐진 형태가 된다. Fig. 7에 대표적인 2차원변형 결과와 Fig. 8에 3차원변형 결과를 예시하고 있다. Fig. 7의 위 두 결과는 에지노드에서 실루엣으로 매핑하는 경우에는 내부 점에 의한 문제가 발생함을 보여 주고 있다. 아래의 결과

는 실루엣에서 에지로 매핑한 결과를 이용한 2차원 변형결과와 3차원 복원 결과 이미지이다.

앞서 다리가 짧은 경우에 2차원 변형을 하게 되면 다음과 같이 다리부분이 늘어나서 이미지에 맞춰지는 긍정적인 효과를 보여준다. 하지만 이 과정에서 원하지 않는 메시의 변형이 일어나게 된다. 이러한 문제를 해결하려면 SMPLify 단계나 후처리 과정을 통하여 조인트 매칭을 할 수 있는 것이 바람직할 것으로 보인다.

3차원 복원은 2차원 복원에 SMPL 신체 모델의 깊이정보를 사용하여 카메라 투영방식으로 이루어진다. 3차원적으로 복원된 구성을 확인하기 위하여 Fig. 8에 복원된 3차원 데이터를 여러 가지 각도에 확인하였다. 3차원 구성자체는 인체자체와 의상을 포함한 사람의 외형차가 크지 않아 상당히 유사하게 복원되었다. 하지만 옆면의 경우 분할이 정교하지 않아 배경을 텍스처로 사용한다던가, 보이지 않는 영역은 투영된 화소를 사용하게 된다. 대칭성이 있는 하의나 상의는 크게 어색하지 않으나 머리 뒷면 등은 매우 어색한 형태로 복원이 됨을 확인할 수 있다. 이 부분은 조건적으로 처리하거나 앞뒷면 두 장을 이미지를 확보하면 해결이 가능할 것으로 보인다.

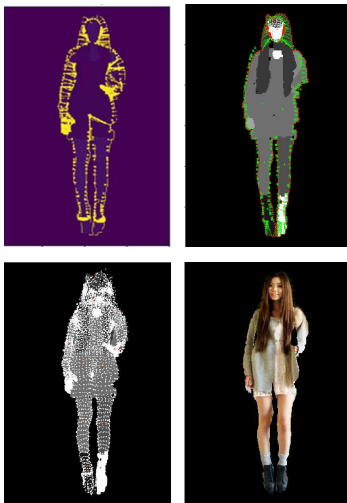


Fig. 7 2D Node Corresponding Points and Deformation Result (Top Left: Mapping from Edge Nodes to Mask Silhouette, Top Right: Mapping from Mask Silhouette to Edge Nodes, Bottom Left: Deformed Mesh, and Bottom Right: Re-projected)

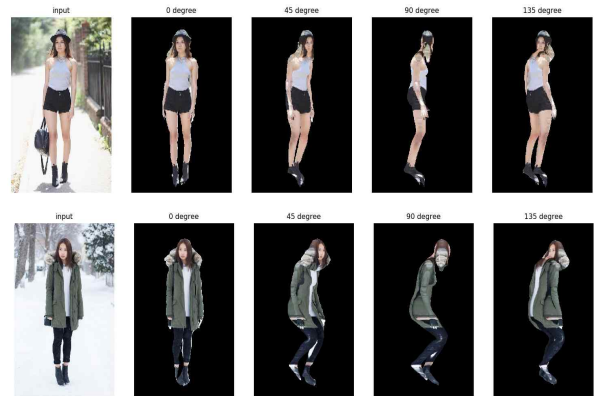


Fig. 8 3D Reconstructed Example (the Original, 0°, 45°, 90°, and 135° Angle Views)

### 3.3 애니메이션 결과

최종적인 애니메이션 결과는 Fig. 9에 제시한 이미지와 링크된 영상을 통하여 확인이 가능하다. 애니메이션 결과는 충분히 흥미롭고, 기능적으로는 원하는 목표를 이루었다고 할 수 있다. 그러나 결과를 객관적으로 정량화하여 평가하기는 어려움이 있다. 몇 가지 성능적인 부분에서 지적할 만한 사항은 다음과 같다. 우선 본 논문에서 개발한 알고리즘의 문제는 아니지만, 이미지 분할 알고리즘이 인물의 실루엣을 확보하는 기술에 한계가 있어, 경계면에 매끄럽고 자연스럽게 만들지 못하다. 단기적으로는 알파 블렌딩 등으로 시각적으로 눈에 뜨지 않게 할 수는 있을 것으로 보이지만 궁극적으로는 물체분할 기술에 개선이 필요하다. 또한 겹치는 영역에서 2차원 매칭점이 자연스럽게 못하면 잘못된 형태의 사람 모델이 복원된다. 가려진 부분의 복원하지 못하면 엉뚱한 신체 부위의 텍스처를 사용하여 렌더링이 이루어지게 된다. 이러한 문제는 현재 텍스처 영역이 가려진 영역인지를 검출하는 기술과 이를 바탕으로 이를 복원하는 기술 두 가지가 필요할 것으로 보인다. 이러한 문제를 해결하는 또 다른 접근 방식으로 여러 장의 사진을 활용하는 것이 있을 수 있다. 고전적인 3차원 카메라 모델을 이용한 3D스캐닝이나 최근 딥러닝 방식의 연구가 있다. 다중 이미지를 사용하기 위해서는 각 이미지를 통하여 얻어지는 정보

를 매핑하고 정렬하는 기술이 선행되어야 할 것이다.

또한 본 논문에서 기초로 하는 3차원 변형 방식은 의상모델이 아니라 피부모델이다. 따라서 느슨한 의상의 자세에 따른 늘어짐 상황은 반영하지 못한다. 의상의 주름이 많은 경우 움직임에 따라 퍼지거나 구겨지는 정도를 반영하지 못한다. 마지막으로 현재 애니메이션에 사용한 걷기 동작파일(BVH) 파일에서 이동위치 등은 사용하지 않고, 걷는 자세만을 활용하고 있다. 이로 인하여 걸음마다 중심위치가 이동하는 대신 매 화면마다 움직임으로서 걷는 모습의 완벽하지 못하다는 문제점도 가지고 있다.



(a) Image #0 (b) Image #195

<https://youtu.be/uuSo7ntjKtg>, [https://youtu.be/aJ\\_E8O\\_Ox1Q](https://youtu.be/aJ_E8O_Ox1Q),



(c) Image #352 (d) Image #539

<https://youtu.be/M4HKcAmUtqk>, <https://youtu.be/HxqY004yYns>

Fig. 9 Fashion Show Walking Animation Results 150th Frame (The Video Clips are available at Youtube Site)

#### 4. 결론

이미지의 대상으로 부터 3차원 모델을 복원하고 이를 애니메이션 하는 기술을 개발하였다. 기능적으로 가능성을 확인하였지만 성능 면에서는 상용적인 수준에 이르려면 여러 요소에서의 개선점을 가지고 있다. 각각 단계별로 해결해야 하는 기술들에 대해서는 결과 분석에서 언급 하였다. 이러한 문제에도 불구하고 본 논문이 유사한 시도를 하는 연구자들에게 의미 있는 해결책을 제시하고 있다고 생각한다. 현재 이러한 부분에 대하여 단기간에 기존에 알려진 방식을 적용하여 해결 가능한 요소들과 난이도가 크고 새로운 해결책이 필요한 것으로 나누어 점진적으로 해결해 나가고 있다.

또한, 서론에서도 언급 한바와 같이 본 논문은 가상착용 결과의 효과적인 시각화하기 위한 기술의 일부로 개발되었다. 즉 단순히 최종적으로 가상착용 결과 3차원 모델을 다각적으로 보여주고 동적으로 움직이는 것을 목적으로 하였지만, 3차원 복원 기술은 다양한 영역에 적용할 수 있을 것으로 판단된다. 예를 들어 이미지 기반 가상 착용기술이 대상 인물의 자세에 따라서 크게 성능이 차이가 나는데, 본 기술을 사용하여 기본 자세로 변환하면 훨씬 좋은 결과가 예상된다.

#### References

Ahn H., (2018a). Online Virtual Try On using Mannequin Cloth Pictures, *Journal of the Korea Industrial Information Systems Research*, 23(6), 29 - 38.

Ahn H., (2018b). Image-based Virtual Try-On System, *Journal of Korean Computer Game Society*, 31(3), 37-45.

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 Image, *European Conference on Computer Vision*, pp. 561-578.

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

Cao, Z., Simon, T., Wei, S. E., and Sheikh, Y. (2017). Realtime Multi-person 2D Pose Estimation using Part Affinity Fields, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7291-7299.

Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. (2017). Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected Crfs, *IEEE Transactions on PAMI*, 40(4), 834-848.

Guo, H., Miao, Z., Zhu, F., Zhang, G., and Li, S. (2014). Automatic Labanotation Generation based on Human Motion Capture Data, *Chinese Conference on Pattern Recognition*, pp. 426-435.

Liu, F., Shen, C., and Lin, G. (2015). Deep Convolutional Neural Fields for Depth Estimation from a Single Image, *Proceedings of the IEEE CVPR*, pp. 5162-5170.

Loper, M., Mahmood, N., Romero, J., Pons-Moll, G., and Black, M. J. (2016). SMPL: A Skinned Multi-person Linear Model, *ACM Transactions on Graphics (TOG)*, 34(6), 248-258.

Parent R. (2012). *Computer Animation, Algorithms and Techniques*, 3rd Edition, Morgan Kaufmann.

Zanfir, M., Popa, A. I., Zanfir, A., and Sminchisescu, C. (2018). Human Appearance Transfer, *Proceedings of the IEEE Conference on Computer Vision and*

*Pattern Recognition*, pp. 5391-5399.

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



**안 희준 (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 : 컴퓨터비전, 컴퓨터 통신, 데이터마이닝



**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 : 컴퓨터비전, 딥러닝, 데이터마이닝