

마네킨 의상사진 기반 온라인 가상 의상 착용

(Online Virtual Try On using Mannequin Cloth Pictures)

안 희 준^{1)*}
(Heejune Ahn)

요 약 본 논문에서 마네킨에 착용된 의상 이미지를 분할하고 사용자의 사진에 입히는 가상 의상 착용 (VTON) 기술을 개발하였다. 의상과 모델의 3차원 정보가 필요하지 않는 2차원 이미지 기반 가상착용연구는 실용적인 가치가 크지만, 연구결과 현재 기술로는 의상 분할 시 가림이나 왜곡에 의한 문제 등 제약사항이 존재한다. 본 연구는 마네킨 의상을 사용함으로써 이러한 어려움을 줄였다는 가정 하에서, 딥러닝 기반 영역분할과 자세추정을 통하여 얻은 결과를 사용자 사진에 입히는 알고리즘을 제안하였다. 기존의 연구 대비 성능 개선을 위하여 사전 자세정보의 신뢰성 검사, 외곽선을 이용한 변형개선, 분할 영역개선 등을 사용하였다. 결과로 시각적으로 만족할 만한 의상착용의 경우가 전체의 50%이상으로 상당히 개선된 결과를 얻었다.

핵심주제어 : 가상착용, 딥러닝, 마네킨, 영상 분할, 이미지 변형

Abstract In this paper, we developed a virtual cloth try-on (VTON) technology that segment the cloth image worn on the mannequin and applies it to the user 's photograph. The two-dimensional image-based virtual wear study which does not require three-dimensional information of cloth and model is of practical value, but the research result shows that there are limitations of of the current technology for the problem of occlusion or distortion. In this study, we proposed an algorithm to apply the results obtained from the DNN- based segmentation and posture estimation to the user 's photograph, assuming that the mannequin cloth reduces the difficulties in this part. In order to improve the performance compared with the existing one, we used the validity check of the pre-attitude information, the improvement of the deformation using the outline, and the improvement of the divided area. As a result, a significantly improved result image of more than 50% was obtained.

Key Words : VTON, Deep Learning, Mannequin, Segmentation, Deformation

1. 서 론

* Corresponding Author : heejune@seoultech.ac.kr

+ 이 논문은 본 연구는 서울과학기술대학교 교내연구비의 지원으로 수행되었습니다(2018-0517).

Manuscript received November 26, 2018 / revised December 10, 2018 / accepted December 16, 2018

1) 서울과학기술대학교 전기정보공학과, 제1저자, 교신저자

2010년 인터넷을 사용한 온라인 상거래의 시장규모가 점점 커지고 있다. 판매 대상의 범위도 오프라인시장의 거의 모든 상품이 가능하게 되었다. 의류 패션 시장의 경우도 매출 규모는 매년

급속히 증가하여, 미국의 경우 2017년 32조 규모이며, 유럽의 경우 16조원규모로 추산된다. 국내의 경우 2017년 추산 약 6000억원으로 전체 온라인시장의 10%에 달하고, 매년 5~10%정도의 증가율을 보인다[1, 2]. 그러나 전자제품 등과 달리 의류는 규격화가 쉽지 않고 형태와 색상, 재질이 매우 다양하며, 개인적인 선호도의 차이가 크기 때문에 온라인에서 사진만으로 구매결정이 어렵다. 이런 이유로 특히 반품이 많이 발생하고 저가 제품위주로 거래가 형성되어있다.

이러한 문제를 완화하기 위하여 최근 가상 의상 착용 (Virtual Try-On: VTON) 기술이 개발되고 있다. 국내에서는 ETRI 등이 이 기술을 개발하여 시연을 한바있으며, 해외에서는 Tri Mirror, Fit Me 등의 회사가 이를 개발하여 상품화하고 개선작업을 진행 중이다[3]. 그러나 게임에 사용되는 3차원 그래픽 렌더링 기술을 사용한 방식으로 의상과 사용자의 3차원 모델정보가 필요하다. 의상과 고객의 3차원정보를 얻기 위해서는 상당한 시간, 돈, 노력이 필요하므로 범용화가 잘 진행되고 있지 못하다.

이러한 한계점을 해결하기 위한 방법으로 제공되는 패션 이미지만을 사용하여 가상착용하려는 연구가 진행되어 왔다[1, 2, 4, 5, 19]. 그러나, 기존연구 결과들은 의상 파싱, 의상의 입힘 등 부분적인 기술을 제한적인 입력상황에서 시험하여, 전 과정을 통한 결과의 성능을 파악하기 어려웠다. 이러한 부분의 궁금증을 갖고 저자들은 연구 [5]에서 최신 의상분할과 의상착용과정을 적용하여 결과를 검토하였다. 그 결과 Fig 1에서 보는 바와 같이, 일부 만족스러운 결과를 보이는 경우도 있으나, 실용화를 하기에 많은 한계점이 있으며, 그 중 가장 어려운 문제로 판단되는 경우는 입력이미지의 의상이 왜곡이나 가리워짐이 심한 경우, 이를 분할하고 자세가 다른 사용자에게 착용하였을 때 결과의상의 품질이 매우 떨어진다는 점이었다.

본 연구에서는 실용성을 고려하여 가리워짐이나 변형이 많지 않은 마네킨 의상을 입력으로 사용하는 경우에 대하여 기존의 알고리즘을 하여

상당한 품질의 개선을 얻을 수 있었다. 본 논문은 우선 제 2절에서 본연구의 근간이 되는 이미지 기반 VTON 시스템 전체 처리흐름과 각 요소 기술에 대하여 설명한다. 또한 실험이 수행된 데이터와 실험 환경에 대하여 기술한다. 제 3절에서는 기존 VTON 시스템의 한계를 극복하기 위하여 입력 의상사진을 마네킨 의상이미지로 제한한 시스템에 대하여 설명하고, 성능 개선을 위하여 추가한 기능요소들과 그 결과들을 제시한다. 또한 실험결과로 나타난 정성적인 결과들을 논의한다. 끝으로 제 4절에서 현 시스템의 성능의 제약을 주는 문제점들을 분석하고 향후 개선할 수 있는 방안에 대하여 제시하면서 논문을 마무리한다.

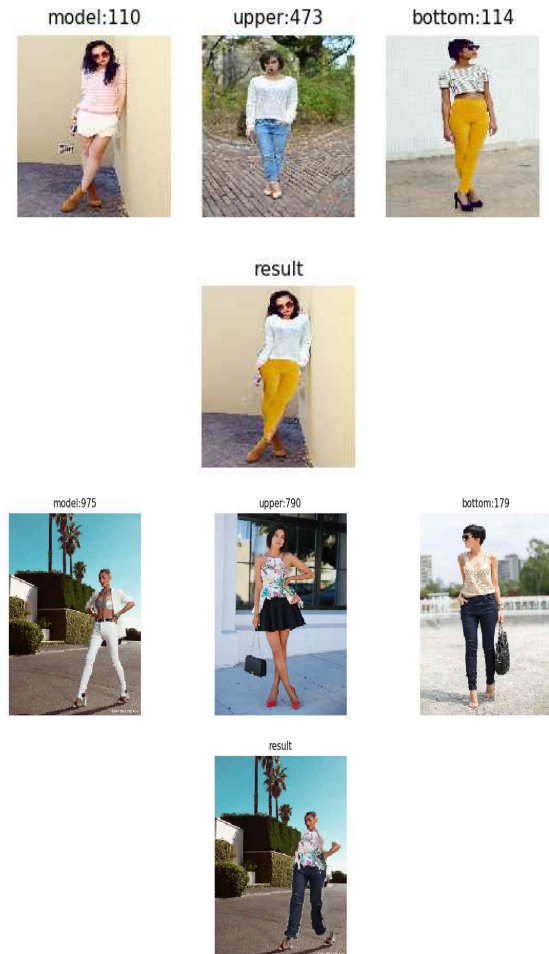




Fig. 1 Model Cloth Based VTON System Results [5] (Top: Successful, Middle and Bottom: Unsuccessful)

2. 이미지 기반 VTON 시스템

본 연구에 사용된 마네킨 기반 VTON시스템의 처리방식은 기존 연구[4]과 큰 흐름에서 차이가 없다. 따라서 본 절에서 이를 간략히 요약하는 형태로 기술하고, 데이터 셋의 차이에 대하여 설명한다.

Fig. 2는 이미지 기반 VTON 시스템의 처리과정을 도식화한 것이다. 전체 응용 시나리오는 크게 의상과 사용자 이미지 등록 단계와 착용단계로 구성된다.

<이미지 등록> 단계에서는 의상이미지를 제한하는 방식으로 의상분할 처리와 자세추정이 수행되고 그 결과 라벨영상과 자세정보를 저장하는 단계이다. <사용자 이미지 등록> 단계에서는 의상이미지와 마찬가지로 사용자 이미지의 영역분할과 자세정보가 추출된다. 이를 위해서는 의상의 종류와 자세에 대한 분류가 필요한데 이는 사용하는 알고리즘에 의존적으로 본 논문에서 사용된 정의는 Table 1과 같다.<가상 착용> 단계에서는 사용자가 자신이 원하는 의상의 이미지와 종류를 선택하고 가상 착용하는 단계이다.

2.1 실험 데이터셋 및 환경

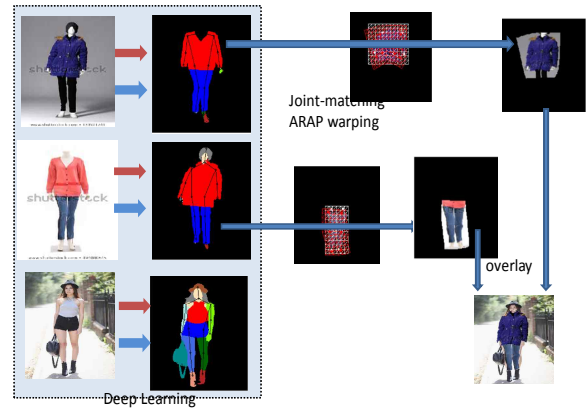


Fig. 2 VTON System Application Scenario Example

본 연구에서 사용된 구현환경은 Python 기반으로, 신경망 네트워크는 Tensorflow 기반으로 구현하였다. 신경망 모델은 요소기술 논문의 저자가 공개하거나, 많은 사용자를 통해서 검증된 코드를 기반으로 부족한 부분은 직접 추가 설계 구현하였다. 트레이닝과 테스트에 사용한 데이터는 논문 [4]에서 제공한 10kdata를 사용하였다. 총 10,003개의 이미지 중 9003개는 학습용 1000개는 시험용으로 사용하였다. 또한 포즈추출을 위해서는 [6]와 [7]에서 사용되는 포즈 모델의 14가지 골격을 사용한다. 또한 3절에서 사용한 마네킨 영상은 인터넷을 통하여 해외 쇼핑몰에서 200개의 임의의 이미지를 확보하였다.

Table 1 Joints and Labels

Joints	Labels
Head top (13)	background (0)
Neck (12)	face (11), sunglasses (3)
Right shoulder (8)	hat (1), hair (2)
Right elbow (7)	upper-clothes (4)
Right wrist (6)	skirt (5), pants (6)
Right hip (2)	dress (7), belt (8)
Right knee (1)	left-shoe(9), right-shoe (10)
Right ankle (0)	left-leg (12), right-leg (13)
Left shoulder (9)	left-arm (14), right-arm (15)
Left elbow (10)	bag (16), scarf (17)
Left wrist (11)	
Left hip (3)	
Left knee (4)	
Left ankle (5)	

2.2 의상 영역분할 알고리즘

영역분할 기술은 컴퓨터 비전의 오랜 연구주제로 많은 연구가 있으나 최근 딥러닝에 방식에 의한 시멘틱 영상 분할기술이 기존의 알고리즘의 성능을 크게 능가하고 있다[8, 9]. 본 연구에서 사용한 영역 분할 기술은 딥러닝기반의 시멘틱 분할기술의 대표적인 모델인 FCN (Fully Convolutional Neural Network) [8]을 사용하였다. FCN 모델의 전단은 Convolutional 방식으로 특징추출을 통하여 Class 분류, 즉 의미 추출을 수행하며, 후단은 Deconvolution 방식으로 영역 분할을 수행한다. 경계영역에서의 영역구분이 잘 되도록 앞단의 계층에서 Bypass 연결을 사용한다. FCN만을 적용한 결과는 화소별로 분류가 일어나게 되므로 독립적인 작은 영역으로 분할이 많이 발생할 수 있다. 이점을 완화하기 위하여 후처리 기술로 그래프 모델에 기반을 둔 Fully Connected Gaussian Edge CRF [10]을 사용하였다. 이 방식은 화소사이를 그래프형태로 연결하고 연결의 인과관계를 화소간의 색상 특성과 거리의 가우시안함수로 모델링하는 방법이다.

10K 데이터 셋을 이용한 학습과 검증 영상에서 의상 분할은 18개의 레이블로 구성되어있고 정답(Ground Truth)영상이 제공된다. Fig. 3는 성공적인 경우와 실패한 경우를 보여준다. 화소별 정확도의 부누적확률 (Negative Cumulative Probability Distribution)을 보여준다. 각각 50%, 80%, 90% 이상의 정확도를 갖는 경우가 90%, 50%, 10%정도를 보였다. 육안으로 확인한 바에 따르면 80%정도의 확률이면 실용적인 가능성이 있다고 판단된다.

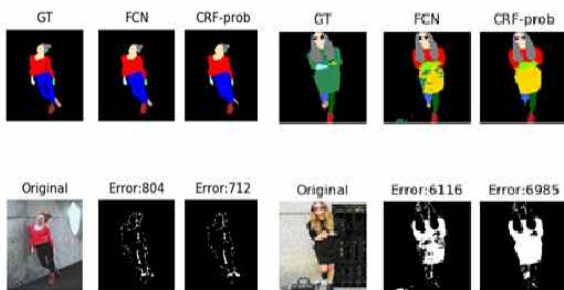


Fig. 3 Parsing Results (Success and Fail) [5]

2.3 자세 추정 네트워크 모델

인간의 활동을 인식하는 것은 매우 응용성이 크기 때문에 자세 추정도 오랫동안 연구되어 왔다. 신체의 일부를 인식하고 이를 통하여 구조를 인식하는 방식이 주로 사용되어 온 방식이나 그 경우의 수가 많아 성능에 제한이 있었다. 최근에 역시 딥러닝에 의한 방식으로 다중 인물의 실시간 자세 인식까지도 어느 정도 가능한 수준의 결과가 발표 되었다. 본 연구는 단계적으로 2016, 2017년에 발표된 ArtTrack [11]과 DeeperCut [12]을 사용한다. 저자들이 네트워크 구현과 기 학습된 파라미터를 제공하고 있어서 재학습 없이 사용하였으며, 모델을 10k데이터셋에 적용하여 자세예측결과를 얻었다.

$$J = \{J_1, \dots, J_P\}, J_i = (x, y) \quad (1)$$

LSP 모델에서 보고된 정확도는 PCK (Percent of Correct Key Points) 82%이상이고, 특히 머리, 어깨, 골반영역에서의 정확도는 95%를 상회하는 것으로 보고되었다. 본 연구에서 사용되는 10k데이터는 자세에 대한 기준에 작성된 정답 (Ground Truth)이 존재하지 않기 때문에, 정량적인 평가는 불가능하다. 따라서 본 논문에서는 다양한 포즈에 대하여 육안으로 확인한 LSP 데이터의 결과이상의 수준인 것을 육안으로 확인하였다. Fig. 4는 자세추정 결과를 Skeleton 다이어그램으로 보여주는데, 서있는 상태에서 사지가 모두 보이는 경우에는 거의 정확한 자세추정을 해내는 반면, 신체가 가려지거나 옆 자세, 안은 자세 등은 정확도가 떨어진다. 기존의 자세 예측이 주로 자세 예측을 통한 행위를 분석하기 위한 것이 라면 본 응용은 좀 더 정확한 예측이 필요한 점이 큰 차이점이라고 할 수 있다.





Fig. 4 Pose Estimation Results

2.4 의상 매칭 및 변형 알고리즘

앞 단계에서 생성된 영역 분할정보와 자세정보를 바탕으로 의상을 입히는 방식은, 모델과 입히려는 의상에 해당하는 골격을 매칭 되도록 드르누이 삼각형 메쉬를 변형하는 방식을 사용한다. 철관으로 모델링하는 TPS(Thin Plate Spline) [13] 방식과 삼각형은 최대한 크기와 형태를 유지하도록 하는 ARAP (As-Rigid-As-Possible) [14]을 구현하여 비교한 결과 TPS를 적용한 결과는 의상착용과 같이 크게 형태가 변하는 경우에는 자연스럽게 못함을 보였다. 따라서 본 연구에서는 ARAP 방식을 적용하였다. 참고로, 의상 핏팅에 대한 연구인 [15]에서도 ARAP 방식을 이용하였으나, 이 논문과 달리 매칭점을 사용자가 직접 입력하는 반자동 방식이 사용되었다. OpenCV와 Scipy의 Piecewise Affine Warp 방식을 사용하였고, ARAP방식은 C++구현을 참고로 Python Numpy를 기반으로 새롭게 구현하였다. 현재 구현은 소요시간이 Mesh의 크기에 따라 제품에 비해하게 증가되는 문제점을 가지고 있다. 현재 10 화소 간격으로 노드를 설정하게 되면 일반적으로 30x30 Mesh정도로 구성이 되는데 이경 경우에 약 3-4초의 동작시간이 소요된다.

3. VTON 알고리즘 개선

3.1 포즈 데이터 검증 알고리즘

화소킨 의상은 모델 의상에 비하여 자세의 가능한 범위가 정의 되어 있다. 따라서 이러한 제약사항을 이용하면 잘못 추정된 포즈영상을 수정은 못하지만 제거할 수는 있다. 다음과 같은 조건을 가정하여 거의 대부분의 잘못된 포즈 정보를 모두 걸러 낼 수 있었다. 조건 1은 좌우 골격이 순서대로 되어 있는지를 조건 2는 위아래로 순서대로 조인트가 구성되어 있는지를 확인하고 있다.

condition 1:

$$(x : J_{pr}(x, y)) - (x : J_{pl}(x, y)) > 0.1H \quad (2-1)$$

condition 2:

$$(y : J_{ph}(x, y)) - (y : J_{pb}(x, y)) > 0.1H \quad (2-2)$$

여기서 H는 마네킨의 키 예측정도

$$H = \max_y(y : J_p(x, y)) - \min(y : J_p(x, y)) \quad (3)$$

이고 pl과 pr는 왼쪽과 오른쪽 쌍으로 존재하는 골격위치로 어깨, 힙, 팔꿈치, 손목, 무릎, 정강이를 의미한다. pt, pb은 상대적으로 적용 되는데, (이마, 목), (목, 왼 어깨), (목 오른 어깨), (오른 어깨 오른 힙) (왼 어깨, 왼 힙), (왼 힙, 왼 무릎) (오른 힙, 오른 무릎), (오른 무릎, 오른 정강이), (왼무릎 왼 정강이)를 사용하였다.

현재 알고리즘으로 검출하지 못하는 경우가 추후에 있더라도 몇 가지 조건을 추가하면 어렵지 않게 검출할 수 있을 것이라고 판단된다. Fig. 5는 실제 검출된 포즈오류의 경우로, 첫 번째는 뒷면으로 인식한 경우, 다리가 꼬인 경우, 그리고 자세가 뭉쳐져 있는 경우를 보여준다.



Fig. 5 Detected Wrong Pose Estimated Cased

3.2 바운더리 정보를 이용한 변형 알고리즘 개선

ARAP 변형은 정확하고 많은 수의 기준점을 사용하면 좀 더 정확한 변형이 가능하다. 기존 연구의 알고리즘에서는 14개의 골격위치를 기반으로 매칭을 하였다. 이렇게 하여 어느 정도 자세 정보는 맞도록 할 수 있었지만, 신체의 두께나 볼륨을 전혀 고려할 수 없었다. 연구 [15]에서는 이를 해결하기 위하여 의상과 모델의 전체 경계 정보를 사용한다. 하지만, 모든 경계정보를 활용하게 되면 의상 자체의 외형이 변형되어 의상의 특성이 사라지게 된다. 기존 연구에서 결과가 잘 나올 수 있었던 것은 사용한 의상이 몸에 달라붙는 형태의 단순한 의상만을 사용하였기 때문이다.

본 연구에서는 상의, 바지, 치마의 3가지 주 의상에 대하여 각기 다른 방식을 적용하였다. 치마의 경우는 보통 체형보다는 의상의 형태를 유지하므로 경계정보를 사용하지 않았다 (향후, 치마의 종류에 따라 다르게 적용할 계획이다.) 상의의 경우는 각각 조인트에 하나의 외각선을 기준점으로 추가하여 활용하여 맞추도록 하였다. 따라서 어느 정도 신체의 굽기 정보가 반영이 되게 하였다. 마지막으로 바지의 경우는 체형정보와 자세에 따른 변형이 충분히 적용이 되도록 조인트와 조인트 중간 위치에서의 경계를 추출하여 기준점으로 사용하였다. 또한 바지의 경우는 외곽과 안쪽을 모두 기준점으로 사용하였다.

경계영역을 영역분할 정보를 활용하면 정확도 높추출 가능한 것을 확인하였다. 여기서 특별히 위치 예측이 어려운 곳이 힙(엉덩이)에 해당하는 영역이다. 힙에 해당하는 영역은 의상의 스타일에 따라서 상의에 위치할 수도 있고 하의에 위치할 수도 있다. 이에 따라 팔의 영역과 구분이 불가능한 경우 발생하였다. 따라서 힙의 경계를 찾을 때는 영역구분 정보대신 어깨와 무릎의 위치로부터 계산을 통해 유추하는 방식을 사용하였다. 즉 Fig. 6과 같이 어깨와 힙 그리고 무릎에 선을 그리고 힙에 수평으로 만나는 점을 사용하였다.

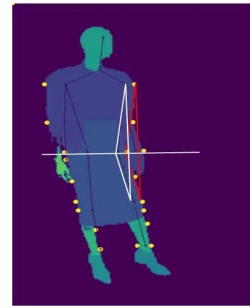


Fig. 6 Estimation Method for the Hip Boundary Location

Fig. 7는 이렇게 하여 얻어진 경계영역을 보여준다. 일반적으로 마네킨 이미지의 경우는 자세 및 체형이 표준적이기 때문에 비슷하면서 정확도가 높은 결과를 얻을 수 있었다. 반면 사람 모델의 경우는 체형에 많은 차이가 있을 수 있는 것을 확인하였다. Fig. 6에서도 대상 모델은 엉덩이와 넓적다리 부위가 유난히 큰 체형이며 이에 따라서 오차가 발생하는 것을 확인하였다. 따라서 체형에 대한 통계적인 조사와 연구가 필요할 것으로 판단된다. 이렇게 해서 얻어진 대표 결과를 Fig. 8에 바지를 입히는 경우와 치마를 입히는 경우로 각각 제시하였다.

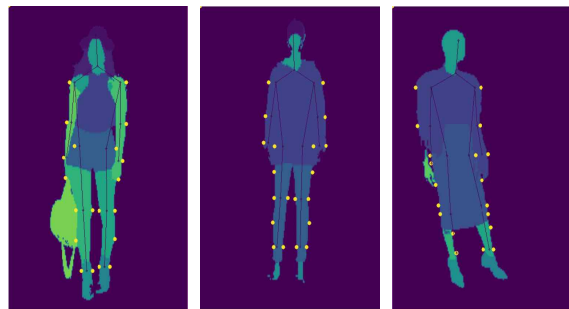


Fig. 7 Boundary Key Points Extracted (left: short clothes, middle: long upper cloth and pants, right: skirt)

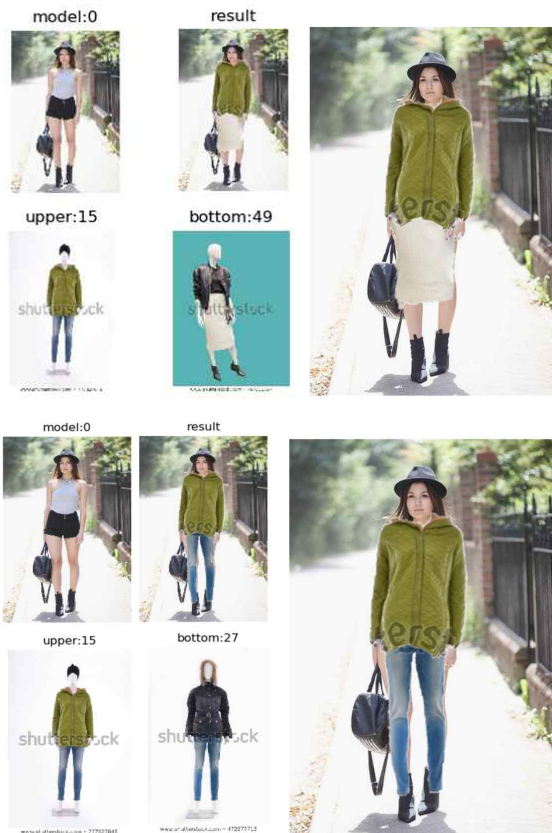


Fig. 8 The Input and Final VTON Image (Enlarged) (Left: Upper and Skirt, Right Upper and Pants)

3.3 의상 경계 영역 개선 및 시각적 개선

DNN 기반 시멘틱 분할로 얻어지는 의상영역 구분 결과는 마네킨 의상의 경우 평균적으로 80-95% 정도의 정확도를 얻을 수 있었다. 하지만 의상 내부에 비하여 경계선상 화소들의 정확도는 상대적으로 낮아, 경계를 기준으로 5화소 범위의 정확도는 평균정확도 대비 10-20%가량 낮은 특성을 보인다. 이는 CNN 방식이 영역분할에 경계정보 보다는 주위의 화소들의 정보를 활용하기 때문으로 추측된다. 따라서 장기적으로는 경계 정보에 민감한 분할 모델을 개발할 필요가 있다.

본 연구에서는 단기적인 방안으로 일차적으로 그래프-컷 (Grab-Cut) [16] 알고리즘과 알파-블렌딩 (Alpha-Blending) [17] 방식 사용하여 개선하는 방식을 사용하였다. 일차로 그래프-컷 알고리

즘을 통하여 대상 의상영역과 이외의 영역의 색상분포정보 모델을 만들고 이를 바탕으로 하여 변화가 큰 화소 경계를 재-검출함으로써 영역마스크를 개선한다. 이후에 수정된 의상영역 마스크를 식 (4)의 알파-블렌딩으로 경계영역의 변화를 완만하게 주면 시각적 측면에서 경계영역의 오류가 눈에 띄지 않아 자연스러운 결과영상을 얻을 수 있다 (Fig. 9).

$$I_{blended}(x,y) = \alpha(x,y)I_{fg}(x,y) + (1-\alpha(x,y))I_{bg}(x,y) \quad (4)$$

여기서 가중치 변수는 해당 의상영역의 마스크를 가우시안 필터로 스무딩 ($\alpha(x,y) = mask(x,y) * N_{\sigma}(x,y)$)하여 얻는다. 실험에는 필터링 범위변수는 $\sigma = 5$ 로 사용하였다. 여기서 입력이미지의 배경이 일반적인 경우로 가정하였다. 만약 특정 단색으로 구성된 경우라면 크로마키 방식을 이용하여 훨씬 더 쉽게 정교한 결과를 얻을 수 있다.



Fig. 9 Boundary Enhancement and Blending Effects (Left: Before, Right: After)

3.4 결과 분석

본 논문에서 제안한 마네킨 의상 입력과 개선된 방식을 사용함으로써, 기존의 방식에 비하여 상대적으로 상당히 실용적가능성이 있는 결과를 얻을 수 있었다. 특히 입력 의상의 왜곡이나 가리워짐의 경우는 확연히 줄었고, 경계매칭을 통하여 부정확한 포즈예측이 있는 경우에도 의상과 모델의 매칭 정도를 높일 수 있었다. 의상의 착용결과에 대해서는 정량적인 기준이 모호하여 이 논문에서는 제시하지 못하나, 자체 시각 평가 결

과 자세와 분할정보가 성공한 입력의 경우 50% 정도는 매우 만족할 만한 최종결과를 (Fig. 8), 30%내외는 약간 어색하지만 형태가 갖추어진, 나머지 20%는 모델의 신체가 제대로 커버되지 않는 등 착용에 실패 (Fig. 10)하였다.

일부 불완전한 경우는 매칭점의 수를 증가하거나 구현 세부알고리즘을 개선함으로써 좀더 나은 결과를 얻을 수 있다. 반면 알고리즘 자체가 가지고 있는 한계도 있다. 예를 들어 Fig. 7에서 입히려는 바지가 상의에 의하여 윗부분이 가려져 있다. 따라서 최종 VTON의 결과 해당영역을 채울 수 있는 의상 정보가 불충분하다. 이러한 점을 해결하기 위해서는 In-Painting 방식이 필수적으로 필요하다. 고전적 영상처리 알고리즘은 입력 영상이나 알고리즘에 정해진 규칙에 의해서만 결과가 나오므로 이를 해결하기가 쉽지 않다. 최근 관심을 끄는 딥러닝의 경우는 모델 안에 기억요소가 포함되어 있어 현재 주어진 일부정보로부터 좀 더 그럴듯한 정보를 생성해내는데 장점이 있다. 최근에 CP-VTON[18]방식의 DNN 기반의 VTON이 제안되었는데 이 결과를 보면 이런 문제를 상당히 해결할 수 있을 것으로 보인다. 단 본 저자들이 실험한 결과 모델이 입고 있는 옷과 입히려는 옷의 형태가 비슷한 경우에 한하여 성공적인 결과를 보이고 그렇지 않은 경우에는 기존의 옷의 형태로 변형하는 문제점을 가지고 있는 것을 확인하였다. 현재 이러한 부분을 해결하기 위한 방안을 연구하고 있다.



Fig. 10 VTON Results with Uncovered Area from Input Clothes

4. 결론

본 논문에서는 의상과 모델의 3차원 정보가 필요하지 않는 이미지기반 VTON기술의 실용화에 대하여 연구하였다. 결과 품질에 입력 의상정보의 완전성이 매우 중요하다는 저자들의 이전 연구 결과를 바탕으로 마네킨 이미지를 사용한 방식을 제안하였고, 수집된 데이터를 바탕으로 VTON을 적용하였다. 또한 마네킨 이미지의 특성을 이용하고 기존 알고리즘을 개선하여 경계점을 이용한 매칭 및 변형, 입력 정보의 자세오류 검출, 경계영역 개선 및 시각적 처리 등을 수행하였다. 그 결과 기존의 모델 영상을 사용한 경우에 비하여 큰 개선을 보였다.

그러나, 실용화에 적용하기에 충분한 수준이라고 보기 어려우며, 세부 알고리즘의 정교화와 함께, 전통적인 이미지 및 그래픽스 방식의 한계점도 노출되었다. 이러한 부분은 입력영상 하나에서 얻어지는 정보로는 해결하기 힘들며, 데이터셋의 통계적인 특성을 반영만 가능할 수도 있을 것이라고 본다. 최근 급격한 발전을 하고 있는 딥러닝 연구는 많은 파라미터를 고려하여 기존에 불가능하던 영역의 결과를 내고 있다 [19-21]. 특히 생성모델(GAN)을 사용한 알고리즘이 그러한 가능성을 보이고 있으나, 아직 원하는 수준의 결과는 보고되지 않았다. 또한 영상의 텍스처나 섬세한 선들이 왜곡 되거나 뭉개지는 현상이 강하기 때문에 현재까지는 본 연구에서 제안하는 바와 같이 하이브리드 한 방식이 효과적이라고 판단된다.

References

[1] Yamaguchi, K., Kiapour, M. H., Ortiz, L. E., and Berg, T. L., "Parsing Clothing in Fashion Photographs," in Proc. of Computer Vision and Pattern Recognition (CVPR), 2012.

[2] Yamaguchi, K., Kiapour, M. H., and Berg, T. L., "Paper Doll Parsing: Retrieving

- Similar Styles to Parse Clothing Items,” in Proc. of International Conference Computer Vision (ICCV), 2013.
- [3] Kim, J., and Forsythe, S., “Adoption of Virtual Try-on Technology for Online Apparel Shopping,” *Journal of Interactive Marketing*, Vol. 22, No. 2, pp. 45-59, 2008.
- [4] Liang, X., Liu, S., Shen, X., Yang, J., Liu, L., Dong, J., and Yan, S., “Deep Human Parsing with Active Template Regression,” *IEEE Tran. on Pattern Analysis and Machine Intelligence*, Vol. 37, No. 12, pp. 2402-2414, 2015.
- [5] Ahn, H., “Image-Based Virtual Try-On System,” *J. of Korean Computer Game Society*, Vol. 31, No 3, pp. 37-45, 2018.
- [6] Johnson S., and Everingham, M., “Clustered Pose and Nonlinear Appearance Models for Human Pose Estimation,” In Proc. of the 21st British Machine Vision Conference (BMVC2010), 2010.
- [7] Andriluka, M., Pishchulin, L., Gehler, P., and Bernt, S., “2D Human Pose Estimation: New Benchmark and State of the Art Analysis,” *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.
- [8] Long, J., Shelhamer, E., and Darrell, T., “Fully Convolutional Networks for Semantic Segmentation,” In Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3431-3440, 2015.
- [9] Vijay, B., Kendall, A., and Cipolla, R., “Segnet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 39, No. 12, pp. 2481-2495, 2017.
- [10] Krähenbühl, P. and Koltun, V., “Efficient Inference in Fully Connected Crfs with Gaussian Edge Potentials,” In *Advances in Neural Information Processing Systems*, 2011.
- [11] Insafutdinov, E., Pishchulin, L., Andres, B., Andriluka, M., and Schiele, B., “DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model,” In *European Conference on Computer Vision (ECCV)*, 2016.
- [12] Insafutdinov, E., Andriluka, M., Pishchulin, L., Tang, S., Levinkov, E., Andres, B., and Schiele, B., “ArtTrack: Articulated Multi-person Tracking in the Wild,” In *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.
- [13] Bookstein, F. L., “Principal Warps: Thin-Plate Splines and the Decomposition of Deformations,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 6, pp. 567-585, 1989.
- [14] Igarashi, T., Moscovich, T., and Hughes, J. F., “As-Rigid-as-possible Shape Manipulation,” *ACM Transactions on Graphics*, Vol. 24, No. 3, pp. 1134-1141, 2005.
- [15] Kim, N., Yoon, J. C., and Lee, I. K., “Image-Based Dress up System,” In *ACM SIGGRAPH’09*, pp. 35, 2009.
- [16] Rother, C., Kolmogorov, V., and Blake, A., “Grabcut: Interactive Foreground Extraction Using Iterated Graph Cuts,” In *ACM Transactions on Graphics (TOG)*, Vol. 23, No. 3, pp. 309-314, 2004.
- [17] Baudisch, P., and Gutwin, C., “Multiblending: Displaying Overlapping Windows Simultaneously without the Drawbacks of Alpha Blending,” In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems ACM*, pp. 367-374, 2004.
- [18] Wang, B., Zheng, H., Liang, X., Chen, Y., Lin, L., and Yang, M., “Toward Characteristic-Preserving Image-Based Virtual Try-On Network,” In Proc. of the

- European Conference on Computer Vision (ECCV), pp. 589-604, 2018.
- [19] Lee, J. H., Kim, B. M., and Shin, Y. S., “Effects of Preprocessing and Feature Extraction on CNN-Based Fire Detection Performance,” J. Korea Industrial Information System Research, Vol. 23, No. 4, pp. 41-53, 2018.
- [20] Kim, D. H. Choi, S. W., and Kwak, S. Y., “Deep Learning Based Fake Face Detection,” J. Korea Industrial Information System Research, Vol. 23, No. 5, pp. 9-17, 2018.
- [21] Ahn, H., “A Study on Compression of Connections in Deep Artificial Neural Networks,” J. Korea Industrial Information System Research, Vol. 22, No. 5. pp. 17-24, 2017.



안 회 준 (Heejune Ahn)

- 정회원
- 2000년 : KAIST 전기정보 공학과 박사
- 1998년 ~ 2002년 : (주) LG전자 차세대단말연구소 선임연구원
- 2002년 ~ 2004년 : (주) Tmax 소프트 책임연구원
- 2004년 ~ 현재 : 서울과학기술대학교 전기정보 공학과 정교수
- 관심분야 : 인터넷 프로토콜, 영상처리, 데이터 마이닝