

# 이미지 입력을 이용한 3차원 아바타의 데이터기반 생성 기법

여영인\*<sup>o</sup> 서혜원\*\* 원광연\*

한국과학기술원 가상현실 연구실\* 충남대학교 컴퓨터그래픽스 연구실\*\*

yyigy@vr.kaist.ac.kr\* hseo@cs.cnu.ac.kr\*\* wohn@vr.kaist.ac.kr\*

## A Data-Driven Approach to 3D Avatar Creation from Photos

Young In Yeo\*, Hyewon Seo\*\*, Kwangyun Wohn\*

Virtual Reality Laboratory at KAIST\*

Computer Graphics and Applications Laboratory at ChungNam National University\*\*

### 요약

본 논문은 인체 스캔 데이터를 예제 데이터베이스로 사용하여 2차원 사진으로부터 3차원 아바타 모델을 생성하는 기술을 제안한다. 직접 기하학적인 변형을 통해 3차원 아바타를 생성하는 기존의 방법들과는 달리, 미리 스캔한 다수의 3차원 인체 형상 모델 데이터베이스를 사용하여 사진에 나타난 실루엣과 가장 일치하는 기본 모델을 변형을 계산함으로써 모델을 획득한다. 본 연구는 예제 기반의 모델 변형 방법이 가지는 사실적인 인체 모델 생성의 장점, 2차원 사진을 이용한 방법이 가지는 텍스처 매핑 및 실루엣 정보 반영의 장점을 모두 가진다. 변형된 모델에 텍스처 매핑을 가하여 얻은 결과 모델에 애니메이션을 적용하여 가상환경에 활용할 수 있다.

### 1. 서론

가상환경에서의 3차원 아바타는 의복 전자상거래 시스템부터 인간 공학적 제품 설계 및 디자인, 온라인 게임 및 애니메이션, 의료 및 건강관련 산업까지 다양한 분야에서 활용되고 있으며 그 분야가 점점 넓어지고 있다. 본 논문은 그 중에서 의복 전자상거래 시스템에 사용되는 3차원 아바타 생성에 초점을 두었다. 오늘날 의복 전자상거래 시스템 구축을 위해 고객의 신체 치수를 반영하는 아바타 생성, 그리고 실시간 의복 시뮬레이션 등의 연구가 활발하게 진행되고 있다[1].

대부분의 3차원 아바타는 디자이너들이 상용 그래픽스 패키지를 사용하여 직접 제작하여 왔다. 이러한 방법은 사실적인 인체 모델을 얻기 어렵고 제작기간이 길다는 단점이 있기에, 자동으로 아바타를 생성하는 방법들이 제안되었다. 초기의 연구 결과들은 대부분 입력된 2차원 사진이나 영상 자료에서 추출한 정보를 이용하여 기본 모델을 기하학적으로 변형하였다 [2][3]. 이러한 방법들은 입력 데이터를 획득하기가 비교적 용이하고, 입력된 사진을 텍스처 이미지로 사용함으로써 유사성이 높은 모델을 얻을 수 있다는 장점이 있다. 하지만 2차원 사진만을 바탕으로 하여 기본 모델을 확대 및 축소하는 방법은 사진의 실루엣에 지나치게 의존하기 때문에, 에러에 민감할뿐만 아니라 기존에 알려진 3차원 아바타 모델이 가지는 보편적인 특성이나 다양성 등을 활용할 수 없다는 단점이 있다.

최근에는 3차원 스캔 장비의 발전과 더불어 3차원 인체 스캐너를 이용한 아바타 생성에 관련된 연구들이 진행되고 있다 [4]. 비교적 짧은 시간에 사용자의 모습과 매우 유사하며 정밀한 3차원 모델을 생성하는 장점이 있지만, 인체 스캐너는 여전히 고가의 장비이며, 의복 전자상거래 시스템에 사용하기 위해서는 스캔 결과물에 복잡도 줄이기와 같은 일련의 후처리를 반드시 거쳐야만 한다.

본 논문의 목표는 입력된 2차원 사진에 나타난 사용자의 모습을 재현하는 동시에, 기측정된 3차원 인체 스캔 모델들로 구축된

데이터베이스로부터의 통계적 추측을 유도함으로써 인체 모델이 지니는 보편적인 특성 역시 잘 반영하는 3차원 아바타를 생성하는 것이다. 생성된 결과 모델은 의복 전자상거래 시스템에 이용되어 사용자가 가상의 의복을 입혀보고 시뮬레이션 함으로써 의복 구매 결정 등에 도움을 줄 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 아바타 생성 관련 연구를 소개하고, 3장부터 5장에는 각각 전체적인 시스템의 개요, 2차원 사진 입력 모듈, 그리고 모델 변형 알고리즘 모듈을 상세히 기술한다. 6장 실험결과와 7장 결론으로 끝을 맺는다.

### 2. 관련연구

이미지 기반의 아바타 모델 생성 기법에 관련된 연구는 주로 얼굴 모델링에 중점이 되어왔다. 예제기반의 얼굴 모델링 기법을 제안한 Blanz와 Vetter[5]는 3차원 얼굴 스캔 데이터를 기반으로 입력된 사진과 유사하게 기본 얼굴 모델을 변형하여 3차원 얼굴을 생성해낸다. 비교적 단순한 얼굴 모델에 비해 신체의 경우에는 복잡한 관절 구조, 모습의 다양성과 더불어 팔, 다리가 신체를 가리는 자기 폐색(Self-occlusion)등 어려움이 많다. 신체 모델링의 경우, 입력된 사진을 바탕으로 기본 모델을 변형하는 연구는 Hilton 외[2] Lee 외[3]의 경우가 있다. 모두 착복한 사용자의 사진에서 추출된 실루엣에 적합하도록 기본 모델을 변형하는 방법이다. 이러한 방법들은 사진의 실루엣 정보만을 이용하기에 데이터기반의 방법보다 3차원 인체 모델의 현실성이 부족하다.

### 3. 시스템 개요

본 논문에서 제안하는 3차원 아바타 생성 방법은 그림 1 과 같다. 사용자는 전면, 측면, 후면의 사진을 입력한다. 시스템은 먼저 전면과 측면 사진을 이용하여 다음의 작업을 실행한다. 사용자가 입력한 2차원 사진에서 실루엣을 추출한 후, 사용자 상호작용(User-interaction)을 통해 특징점(Feature points)을 입력 받는다. 3차원 기본 모델은 전면, 측면 사진에 투영된 후 사진과의

실루엣 비교를 통해 에러를 계산한다. 계산된 에러를 바탕으로 인체 스캔 데이터베이스에서 주성분 분석방법을 이용해 데이터를 검색하고 이에 맞게 기본 모델을 변형하는 작업을 에러가 충분히 작을 때까지 반복한다. 최종적으로 변형된 모델은 전면 사진에 투영한 후 텍스처 좌표를 생성 및 저장한다. 다음으로 후면, 측면 사진을 이용하여 위의 작업을 반복하여 후면 사진에 기본 모델을 투영하여 텍스처 좌표를 생성 및 저장한다. 이와 같은 과정을 통해 기본 모델은 전면과 후면의 사진들로부터 텍스처가 입혀지며, 추가적인 작업 없이 애니메이션이 가능하다.

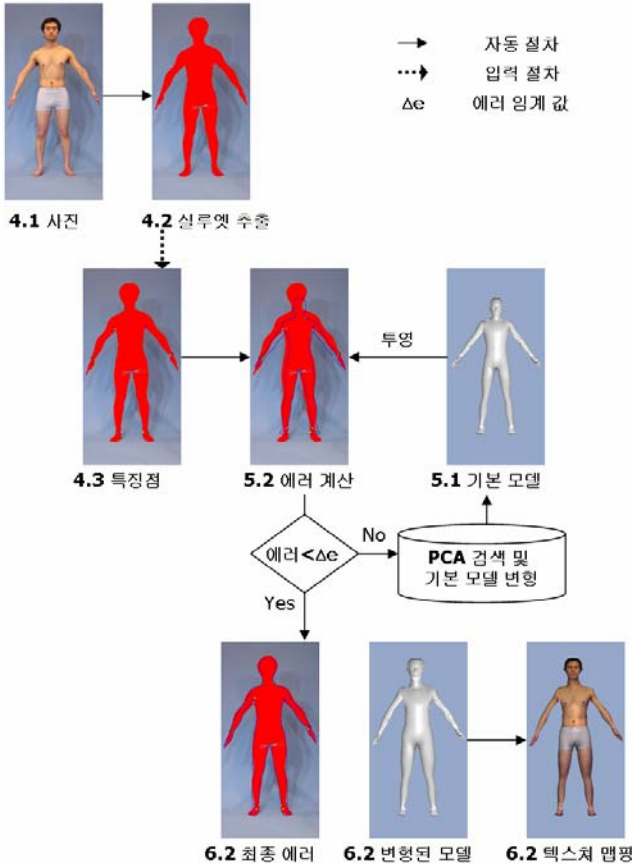


그림 1 전체 작업 구성도

#### 4. 사진 입력 모듈

##### 4.1 사진 촬영

피실험자는 속옷만 착용하고 양팔과 다리를 벌린 자세에서 파란색 배경 앞에서 사진 촬영에 들어간다. 본 논문에서는 전면, 측면, 후면 세 장의 사진이 입력으로 사용되며, 한 대의 카메라만을 이용하여 피실험자가 방향을 90°씩 변경하며 촬영하였다.

##### 4.2 실루엣 추출

다양한 실루엣 추출 방법 중 본 논문에서는 색상 값을 이용하여 2차원 사진에서 인체 실루엣을 자동으로 추출한다. 색상 값이  $(R, G, B)$  인 픽셀이 파란색의 배경으로 분류되는 경우는 다음의 세가지 조건을 모두 만족하는 경우이다. 첫째,  $B$  값이 가장 커야 하며, 둘째, 일정 임계치 이상이어야 하며, 셋째, 가장 작은 색상과 일정 값 이상 차이가 나야 한다. 본 연구에서는  $B$  가 100 보다 크고, 가장 작은 값의 차이가 20 이상이 되도록 설정하였다. 마지막 조건은 신체와 속옷에 나타나는 밝은 부분을 배경으로 분류하는 것을 막기 위해서 필요하다.

#### 4.3 특징점 입력

앞 단계에서 구한 실루엣으로부터 자동으로 특징 점을 찾는 데에는 한계가 있다. 머리끝, 발끝, 손끝과 같은 지점만으로는 아바타의 관절 정보를 얻어내기 힘들다. 그러므로, 3차원 아바타 생성에 필수적인 관절 정보를 유추하기 위한 특징점들은 목, 손목, 발목 등에서 간단한 사용자의 상호작용을 통해 입력된다. 본 논문에서는 전면/후면 사진의 경우 13개, 측면 사진의 경우 10개의 특징점을 사용하였다.

### 5. 기본 모델 변형 알고리즘

#### 5.1 3차원 인체 모델 공간 구성과 탐색

Seo 외 다수는 여러 인체의 3차원 스캔 데이터들을 예제로 사용한 인체 모델 공간을 제안하였는데[6], 본 논문에서는 이 인체 공간을 탐색하여 이미지 공간에서 정의된 특색점 및 실루엣 에러(5.2절 참조)를 최소화하는 매개변수를 찾아냄으로써 형상을 획득한다. 하나의 3차원 인체 스캔 데이터는 기본 모델을 이루고 있는 33개의 관절 각각에 적용된 크기(scale), 회전(rotation), 그리고 위치 이동(translation)을 나타내는 관절 정보(joints parameters)와 기본 메쉬를 구성하는 1000개 이상의 정점들에 대한 x, y, z 축 방향으로의 변위 정보(displacement map)로 표현된다. 이렇게 크기가 방대한 다차원 공간은 효율적인 검색이 불가능하므로, 주성분분석(Principal Component Analysis)을 이용하여 차원을 현저하게 줄이고 30개의 주성분(Principal Component)만으로 원 데이터를 표현한다. 주성분분석은 서로 연관이 있는 변수들이 관측되었을 때, 이 변수들이 가지고 있는 정보들을 최대한 확보하는 적은 수의 새로운 변수들을 생성하는 방법이다. 3차원 인체 스캔 데이터는 관절 정보와 변위 정보가 사람의 형태라는 공통점을 기반으로 하기에 주성분분석이 용이하다. 이렇게 구성된 주성분 공간에서 임의의 인체 형상은 다음과 같이 두 개의 선형 모델로 표현된다.

$$\mathbf{j} = \bar{\mathbf{j}} + \mathbf{P}_j \mathbf{b}_j, \quad \mathbf{d} = \bar{\mathbf{d}} + \mathbf{P}_d \mathbf{b}_d$$

$\bar{\mathbf{j}}$  와  $\bar{\mathbf{d}}$  는 각각 관절 정보와 정점의 변위 정보에 대한 평균 벡터이며,  $\mathbf{P}_j$  와  $\mathbf{P}_d$  는 서로 직교하는 주성분들의 집합이고,  $\mathbf{b}_j$  와  $\mathbf{b}_d$  는 계수들의 집합이다.

그러므로, 5.2절에 정의한 이미지 공간 상에서의 에러가 최소가 되도록  $(\mathbf{b}_j$  와  $\mathbf{b}_d$  들의 집합으로 이루어지는) 인체 모델 공간을 탐색하여 이들을 기본 모델의 변형에 사용함으로써 아바타 모델의 형상을 획득하였다. 이 탐색에는 다차원 함수의 최적 해를 구하기 위해 Powell 외 다수가 제안한 Direction Set 방법[6] 을 사용하였다. 이 방법은 함수의 미분값 계산을 요구하지 않고 함수값만을 직접 사용하므로 본 논문의 경우와 같이 미분 계산이 어려운 다차원 함수를 다루는 경우에 적합하다.

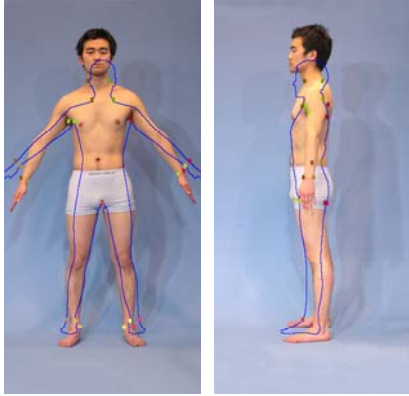
#### 5.2 에러 계산

입력된 사진과 기본 모델을 이미지에 투영하여 추출된 두 실루엣을 비교하여 그 유사 정도를 정량적으로 측정할 수 있도록 에러를 크게 다음의 두 가지로 정의하였다. 첫째는 대응하는 특징점간의 거리의 제곱 값이며, 둘째는 실루엣의 불일치로 인해 생겨나는 픽셀 영역의 비율이다. 다음에 이들 에러의 정의에 대해 보다 상세히 기술한다.

##### 5.2.1 특징점간의 거리

$$E_d = \sum_{i=1}^n dist^2(P(F_{T,i}), F_{D,i})$$

$F_{D,i}$  는 사진에서의  $i$  번째 특징점 좌표이고  $F_{T,i}$  는 3차원 기본모델의  $i$  번째 특징점 좌표이다.  $P:R^3 \rightarrow R^2$  를 이용하여 3차원 기본 모델을 2차원 이미지에 직교 투영하여 해당 특징점 좌표를 구한다. 특징점 정보는 3차원 아바타의 관절정보, 즉 크기, 회전, 위치를 유추하는 데 유용하게 사용된다. 해당 특징점들을 그림 2에서 각각 다른 색으로 표현하였다.



(a) 전면 사진 에러 (b) 측면 사진 에러

그림 2 이미지 공간에서의 에러 계산

### 5.2.2 불일치 영역 비율

$$E_a = \frac{\sum (T(i, j) \cdot \bar{D}(i, j))}{\sum T(i, j)} + \frac{\sum (\bar{T}(i, j) \cdot D(i, j))}{\sum D(i, j)}$$

$T(i, j)$ ,  $\bar{T}(i, j)$  는 기본 모델이 투영된 이미지에서  $(I, j)$  좌표의 픽셀이 기본 모델의 내부, 외부를 의미하는 판별 값이다.  $D(i, j)$ ,  $\bar{D}(i, j)$  은 입력된 사진에서  $(I, j)$  좌표의 픽셀이 피실험자의 신체부분인지 배경부분인지 판별하는 값이다. 실루엣이 서로 겹치지 않는 부분의 비율은 신체의 전체적인 모습을 결정하는데 중요하다.

### 5.2.3 에러 통합

앞의 두 가지 에러에 다음과 같이 가중치를 두어 더한 값을 최종 에러로 사용하였다.

$$E = \alpha E_d + (1 - \alpha) E_a$$

초기에는 관절 매개변수를 찾기 위해  $\alpha = 1$  로 설정한다. 특징점간의 거리를 줄이기 위해 3차원 인체 데이터베이스에서 주요 성분 분석을 통해 에러를 최소화하는 관절 정보를 찾는다. 다음으로  $\alpha = 0.3$  로 설정하여 불일치 영역을 최소화하는 변위정보를 찾는다.

## 6. 실험 및 결과

### 6.1 데이터 집합

본 연구에서는 60개의 여성 인체 스캔과 40개의 남성 인체 스캔 데이터를 사용하였다. 이들 3차원 스캔은 Tecmath™ 와 Cyberwear™ 레이저 스캐너를 사용하여 획득되었다. 스캔 자세는 2차원 사진 촬영 자세와 유사하고 스캔 대상자는 속옷만 착용하였으며, 결과 데이터에서 얼굴 부분은 제거하였다.

### 6.2 모델 복원

다양한 입력된 사진을 이용하여 본 논문의 내용을 실험 하였으며 그림 3에 그 중 하나의 결과를 보였다. 실험은 펜티엄 IV 2.6 GHz, 512MB 에서 실행하였으며 주성분분석 검색 시간이 5분 정도 소요되었다. 텍스처 매핑은 변형된 모델의 모든 정점을 전/후면 사진에 투영하여 텍스처 좌표를 구한다. 해당 정점의

법선벡터와 뷰 포인트와의 차이를 기준으로 전면 또는 후면 사진의 텍스처로 매핑한다. 또한 변형된 모델은 관절 정보를 업데이트 함으로써 애니메이션 가능하다. 표 1은 모델 변형에 따른 실루엣 에러 감소를 보인다.

표 1 실루엣 에러

	초기 에러	2 개 관절 주성분 검색	2 개 변위 주성분 검색
$E_d$ (평균)	전면: 323, 측면: 398	145, 109	81, 100
$E_a$ (%)	전면: 56, 측면: 46	29, 30	22, 27



그림 3 실험 결과: 변형된 기본 모델과 애니메이션 적용 결과

## 7. 결론 및 향후 연구

본 논문에서는 세 장의 사진을 이용하여 3차원 아바타 모델을 생성하는 기법을 제안하였다. 전처리된 3차원 인체 스캔 데이터들에 대해 주성분분석을 수행하여 얻은 인체 공간을 검색하여, 입력된 사진의 실루엣에 일치하도록 기본 모델을 변형하였다. 3차원 인체 모델을 직접 수정하지 않고 에러 값을 최소화하는 방향으로 자동 변형이 가능하게 하였다.

향후 연구에서는 전면, 측면, 후면과 같은 정해진 위치가 아닌, 임의의 시점에서 촬영한 사진으로부터 3차원 아바타 생성이 가능하도록 할 계획이다. 또한, 임의의 자세와 배경을 바탕으로 촬영한 사진 입력에 대해서도 동작 가능하도록 할 계획이다.

## 8. 참고문헌

- [1] B. Allen, B. Curless, Z Popovic. "The space of human body shapes: reconstruction and parameterization from range scans", Proc. SIGGRAPH '03, pp. 587-594, Addison-Wesley, 2003.
- [2] B. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces", Proc. SIGGRAPH '99, Addison-Wesley, pp. 187-194, 1999.
- [3] F. Cordier, H. Seo and N. Magnenat-Thalmann, "Made-to-Measure Technologies for Online Clothing Store", IEEE Computer Graphics and Application, pp. 38-48, January/February 2003.
- [4] A. Hilton, J. Starck and G. Collins, "3D Shape Capture to Animated Models", Proc. First International Symposium on 3D Data Processing, Visualization, and Transmission (3DPVT), IEEE Press, Padova, Italy, pp. 246-255, 2002.
- [5] W. Lee, J. Gu, and N. Magnenat-Thalmann, "Generating Animatable 3D Virtual Humans from Photographs", Computer Graphics Forum, vol. 19, no. 3, Proc. Eurographics 2000 Interlaken, Switzerland, pp. 1-10, August, 2000.
- [6] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. Numerical Recipes in C, The art of scientific computing. Cambridge University Press, 1988.
- [7] H. Seo, N. Magnenat-Thalmann, "An Automatic Modeling of Human Bodies from Sizing Parameters", ACM SIGGRAPH Symposium on Interactive 3D Graphics, ACM Press, pp. 19-26, 2003.