

워핑 기법에 의한 얼굴의 포즈 합성에 관한 연구

오승택[†], 서준원^{**}, 전병환^{***}

[†] 공주대학교 컴퓨터공학과, ^{**}(주)모리아테크놀로지, ^{***}공주대학교 정보통신공학부

e-mail : bhjun@kongju.ac.kr

A Study on the Synthesis of Facial Poses based on Warping

Seung Taek Oh[†], Joon Won Seo^{**}, Byung Hwan Jun^{***}

[†]Dept. of Computer Engineering, Kongju National University,

^{**}Moriah Technology, Inc.,

^{***}Division of Information and Communication Engineering, Kongju National University

요약

본 논문에서는 사실적인 아바타(avata) 구현의 핵심이라 할 수 있는 입체적인 얼굴 표현을 위해, 기하학적인 정보를 사용하지 않고 중첩 메쉬를 허용하는 개선된 메쉬 워프 알고리즘(mesh warp alg)을 이용하여 IBR(Image Based Rendering)을 구현하는 방법을 제안한다. 3차원 모델을 대신하기 위해 인물의 정면, 좌우 반측면, 좌우 측면의 얼굴 영상들에 대해 작성된 메쉬를 사용한다. 합성하고자 하는 정면 얼굴 영상에 대해서는 정면 메쉬만을 작성하고, 반측면이나 측면 메쉬는 표준 메쉬를 근거로 만든다. 얼굴 포즈 합성의 성능을 평가하기 위해, 얼굴을 수평으로 회전하는 실제 포즈 영상과 합성된 포에 대해 주요 특징점 들을 정규화 한 위치 오차를 측정 한 결과, 평균적으로 양 눈의 중심에서 입의 리에 대해 약 5%의 위치 오차만이 발생한 것으로 나타났다.

1. 서론

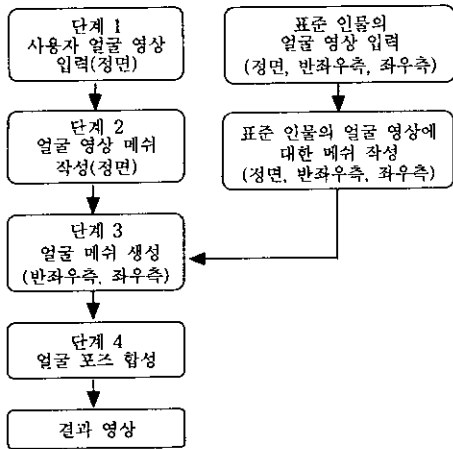
최근에는 3차원 컴퓨터 그래픽 기법의 단점을 보완하기 위해 IBR(Image Based Rendering) 기법이 활발하게 연구되고 있다. IBR 기법은 컴퓨터 그래픽으로 구현하려는 대상의 2차원 영상을 이용함으로써 전통적인 3차원 그래픽 기법의 모델링 비용과 시간을 절감하고, 실제계의 색상과 질감을 그대로 구현하려는 시도라고 볼 수 있다[1]. IBR 기법은 실제계에 존재하는 물체를 적은 비용과 시간을 투입하여 사실적으로 표현하고자 할 경우 적당한 방법이다[2,3]. 또한, 컴퓨터 픽스로 구현해야 할 대상이 매우 많으며, 실시간 처리가 필요한 가상현실 구현에 적절한 방법이라고 할 수 있다. 특히 가상현실에 사용되는 아바타 구현에 IBR 기법을 적용할 경우, 사용자가 자신의 모습을 찍은 실제 영상을 이용하여 손쉽게 사실적이고 독자적인 아바타 제작이 가능함으로써 가

상현실에 대한 참여와 몰입도를 높일 수 있다[4, 5].

본 논문에서는 아바타 구현에 있어서 핵심적인 부분이라고 할 수 있는 다양한 포즈의 얼굴 영상을 합성하는 방법을 제안한다. 본 논문에서 제안하는 방법은 3차원 기하학적 정보를 사용하지 않고 사용자의 얼굴 영상만을 사용하여 원하는 포즈의 얼굴 영상을 합성하는 방법이며, 이로 인해 3차원 모델 구축에 소요되는 비용과 시간을 절감하고, 일반적인 개인용 컴퓨터 환경에서 구현이 가능하며, 비교적 빠른 처리시간을 유지하도록 할 수 있다.

이를 위해서 본 논문에서는 얼굴 포즈 상태를 정의하고, 원하는 포즈에 따라 얼굴을 합성한다. 얼굴 합성에 사용되는 영상은 실제 인물의 정면 얼굴 영상 한 장을 이용한다. 얼굴 합성에 사용되는 알고리즘은 기존의 메쉬 워프 알고리즘을 얼굴 포즈 변화에 적용할 수 있도록 개선하여 적용한다. [그림 1]은 본 논문에서 제안하는 얼굴 포즈 합성의 흐름도이다.

[†] BK21 대전·충남 정보통신인력양성사업단 연구조교(RA) 수혜



[그림 1] 얼굴 포즈 합성 흐름도

2. 얼굴포즈합성

워핑(warping)이란 정해진 규칙에 따라 입력 영상을 재출하는 작업으로, 입력 영상에 설정된 제어점이나 제어 메쉬 사용하여 영상의 화소를 이동시켜 원하는 목적 영상으로 변하는 것을 말한다.

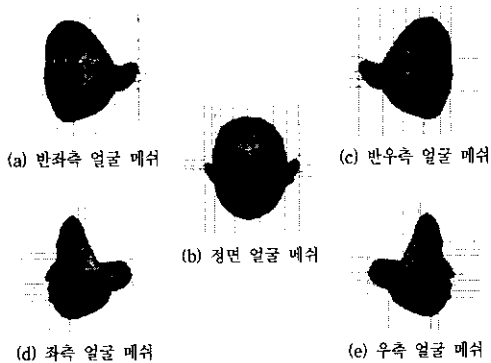
본 논문에서는 메쉬 점의 중첩을 허용하지 않는 기존의 메쉬 알고리즘(mesh warp algorithm)[7]을 얼굴의 회전 변에 의해 이웃한 메쉬 점들이 중첩되는 경우에도 수행될 수 있도록 개선하였다[6].

2.1 얼굴포즈유형정의

인물의 얼굴 포즈는 다양한 유형을 가질 수 있으나, 얼굴 포즈를 회전으로 정의하면 크게 세 가지 형태를 가진다. 긍정 회전(yes-rotation)은 고개를 끄덕이듯이 수직으로 회전하는 것을 의미하고, 부정 회전(no-rotation)은 고개를 가로짓듯이 얼굴을 수평으로 회전시키는 회전을 말하며, 글세 회전(well-rotation)은 고개를 가우뚱하듯이 기울이는 회전을 말한다.

2.2 워핑에 의한 얼굴포즈합성

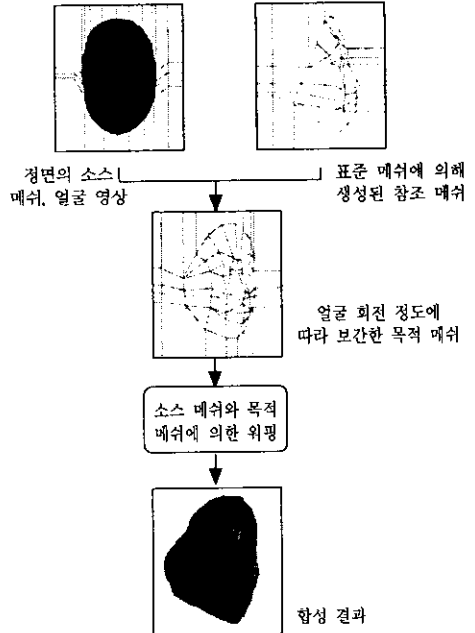
워핑에 의해 부정 회전에 해당하는 얼굴 포즈를 합성하기 위해서는 표준 인물에 대해 [그림 2]와 같이 정의된 포즈 얼굴영상에 메쉬를 설정해야 한다.



[그림 2] 각 방향에 따른 얼굴 제어 메쉬의 설정

이때 정면 메쉬는 소스 메쉬로 사용되고 나머지 메쉬들은 목적 메쉬 생성을 위한 참조 메쉬로서 사용된다. 원하는 방향의 목적 메쉬는 소스 메쉬와 목적 메쉬를 이용하여 보간법으로 생성한다.

[그림 3]은 계산된 목적 메쉬를 이용하여 원하는 포즈의 얼굴을 합성하는 방법을 도식적으로 보여준다.



[그림 3] 계산된 목적 메쉬를 이용한 원하는 얼굴 포즈의 합성

3. 실험 및 분석

3.1 실험환경

실험 영상은 CCD 카메라를 이용하여 연속적인 포즈 변화를 AVI 포맷의 동영상으로 수집하고 필요에 따라 동영상의 프레임들 BMP 포맷으로 추출하여 사용한다. 추출된 영상에 대해서 머리카락과 목 이하 부분을 제외시키고 얼굴 부분만을 변형에 이용할 수 있도록 한다. 이때 필요에 따라 귀도 제거될 수 있다.

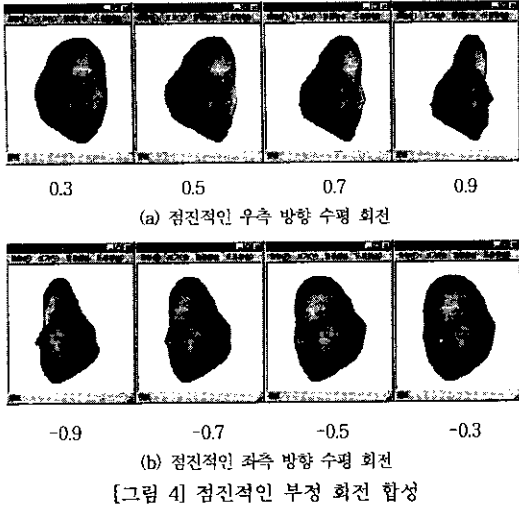
3.2 실험내용

(1) 얼굴 포즈의 합성에

[그림 4]는 회전 각도의 구간 $-90^{\circ} \sim 90^{\circ}$ 를 정규화된 구간 $-1.0 \sim 1.0$ 으로 표현할 때, 점진적으로 부정 회전을 하고 있는 얼굴 포즈의 합성 결과를 보여준다. 여기서, 회전이 심할수록 형태의 왜곡이 점점 심해지는 것을 알 수 있는데, 이는 회전 각도가 커질수록 부정확한 깊이 정보가 차지하는 비율이 점점 높아지기 때문이다.

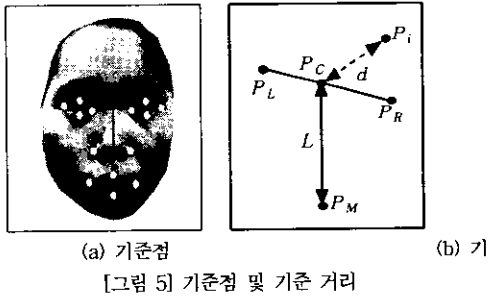
(2) 원 영상과 합성 영상의 비교

인물의 실제 얼굴을 어느 정도 사실적으로 합성할 수 있는지를 평가하기 위해서 본 합성 시스템을 이용하여 합성한 영상과 동일한 포즈의 실제 인물의 영상을 비교한다. 비교 기준은 얼굴의 구성요소인 눈, 코, 입에 설정한 기준 점들 간의 위치 오차를 계산한다.



[그림 4] 점진적인 부정 회전 합성

[그림 5]는 실제 영상에 설정한 비교 기준점과 기준 거리를 보여준다.



[그림 5] 기준점 및 기준 거리

비교 기준점은, [그림 5]의 (a)와 같이, 눈과 입의 경우 각각 해당 영역의 양쪽 수평 끝 두 점과 이 점들을 연결하는 선분의 1/2 지점에 대해 수직으로 두 점을 설정하며, 코의 경우 코의 하단 양 끝 두 점을 설정, 총 14점을 설정한다. [그림 5]의 (b)는 합성영상에서의 기준 거리 L 과 d 를 나타내고, 이에 대응하는 실제 영상에서의 기준 거리는 각각 L' 과 d' 이다.

합성 영상과 실제 영상의 크기가 다를 수 있으므로, 먼저 합성 영상의 각 기준점의 상대적인 위치를 다음과 같이 나타낸다. 즉, 양쪽 눈의 중심을 연결한 직선과 입의 중심점에서 시작하는 수직선이 교차하는 점 P_C 에서 입의 중심점 P_M 까지의 수직 거리 L 에 대한 P_C 에서 각 기준점까지의 거리 d 의 비율로 나타낸다. 마찬가지로 실제 영상에 대해서도 그 비율을 구하고 이들 간의 차이를 모두 평균한 값이 평균 위치 오차 D 이다.

$$D = \left(\sum_{i=1}^{N-1} |d_i/L - d'_i/L'| \right) / N \quad , \quad (N: \text{기준점의 수})$$

[표 1]은 부정회전 정도에 따라 합성한 영상과 실제 영상에 대한 평균 위치 오차를 보여준다.

[표 1] 실제 및 합성 영상에 대한 기준점의 평균 위치 오차

회전 정도	평균 오차(D)	회전 정도	평균 오차(D)
0.1	0.024373	-0.1	0.026789
0.2	0.051771	-0.2	0.020094
0.3	0.054247	-0.3	0.111002
0.4	0.031487	-0.4	0.030525
0.5	0.037015	-0.5	0.018505
0.6	0.034623	-0.6	0.014460
0.7	0.020420	-0.7	0.028245
0.8	0.135441	-0.8	0.128873
0.9	0.050523	-0.9	0.076134
1	0.124846	-1	0.127513
평균 위치 오차		0.051476	

결과적으로, 측정된 평균 위치 오차 D 는, 양 눈의 중심에서 입의 중심까지의 수직 거리에 대해 약 5%로 나타났습니다.

4. 결론

본 논문에서는 기하학적인 3차원 모델을 사용하지 않는 2차원 메쉬 워핑 방법에 의한 얼굴 포즈 합성 방법을 제안하였다. 이를 위해 중첩 메쉬를 허용하는 개선된 메쉬 워프 알고리즘을 사용하였으며, 부정회전에 의한 다양한 포즈에 대해 얼굴을 합성하여 실제 포즈 영상과 비교함으로써 효과적으로 포즈를 합성할 수 있음을 보였다.

앞으로 본 시스템의 개선을 위해서는 다양한 얼굴 포즈와 인물의 표정을 합성할 수 있도록 해야 하며, 회전 정도가 커짐에 따라 발생하는 얼굴 형태의 왜곡에 대한 해결이 필요하다.

참고문헌

- [1] Paul Debevec, "Introduction-What is Image Based Modeling and Rendering", Siggraph98 Course Note, Siggraph, pp.12-14, 1998.
- [2] 최재정, 신영길, "볼륨 데이터의 효율적인 영상 기반 렌더링", 한국정보과학회 논문지(A), 제 26권 3호, pp.261-270, 1999.
- [3] 박동규, 박상현, 이훈희, 조한규, "이미지 합성을 활용한 자동 군중생성 기법", 한국정보과학회 학술발표집, 제 24권 2호, pp.679-682, 1997.
- [4] 공광식, 김창현, "한국어를 말하는 버추얼 캐릭터의 입술 움직임과 표정의 동기화", HCI, pp.599-604, 2000.
- [5] 강명진, 공광식, 김창현, "3차원 개인 얼굴 에이전트의 생성 및 한글 입술 애니메이션의 구현", HCI, pp.109-113, 1999.
- [6] 서준원, "개선된 메쉬 워프 알고리즘에 의한 얼굴 포즈 합성", 공주대학교 대학원 석사학위논문, 2000.12.
- [7] Randy Crane, "Simplified approach to Image Process", Prentice-Hall PTR, 1997.
- [8] Rafael C. Gonzalez and Richard E. Woods, "Digital Image Processing", Addison-Wesley Pub., 1992.
- [9] Richard S. Wright Jr. and Michael Sweet, "OpenGL super bible", The Waite Group, 1996.
- [10] 공광식, 김창현, "입술 애니메이션을 위한 한글 발음의 동시 조음 모델", 정보과학회 논문지, 제 26권 9호, pp.1031-1041, 1999. 9.
- [11] 강명진, 김창현, "얼굴 메쉬와 이미지의 동적 매칭을 이용한 개인 아바타의 3차원 얼굴 합성", 한국정보과학회 가을학술 발표논문지, pp.600-602, 1998.