얼굴 애니메이션을 위한 직관적인 유사 고유 얼굴 모델

김익재°, 고형석
서울대학교 그래픽스 및 미디어랩
{ijkim,ko}@graphics.snu.ac.kr

Intuitive Quasi-Eigenfaces for Facial Animation

Ig-Jae Kim° and Hyeong-Seok Ko
Graphics & Media Lab, Seoul National Univ.

요 약

블랜드 쉐입 기반 얼굴 애니메이션을 위해 기저 모델(Expression basis)을 생성하는 방법을 크게 두 가지로 구분하면, 애니메이터가 직접 모델링을 하여 생성하는 방법과 통계적 방법에 기초하여 모델링하는 방법이 있다. 그 중 애니메이터에 의한 수동 모델링 방법으로 생성된 기저 모델은 직관적으로 표정을 인식할 수 있다는 장점으로 인해 전통적인 키프레임 제어가 가능하다. 하지만, 표정 공간(Expression Space)의 일부분만을 커버하기 때문에 모션데이터로부터의 재복원 과정에서 많은 오차를 가지게 된다. 반면, 통계적 방법을 기반으로 한 기저 모델 생성 방법은 거의 모든 표정공간을 커버하는 고유 얼굴 모델(Eigen Faces)을 생성하므로 재복원 과정에서 최소의 오차를 가지지만, 시각적으로 직관적이지 않은 표정 모델을 만들어 낸다. 따라서 본 논문에서는 수동으로 생성한 기저모델을 유사 고유 얼굴 모델(Quasi-Eigen Faces)로 변형하는 방법을 제시하고자 한다. 결과로 생성되는 기저모델은 시각적으로 직관적인 얼굴 표정을 유지하면서도 통계적 방법에 의한 얼굴표정 공간의 커버 영역과 유사하도록 확장할 수 있다.

1. Introduction

움직이는 캐릭터의 얼굴 표정 애니메이션은 스토리를 전달하는 중요한 역할을 한다. 따라서, 애니메이션 스튜디오에서는 사실적인 표현이 가능한 얼굴 표정을 만들어내는 것이 가장 중요한 기술중의 하나로 여겨진다. 그럼에도 불구하고, 현재까지 얼굴 표정을 생성하는 방법에 있어서 표준화된 절차가 없이 노동 집약적인 작업에서 부터 최신 기술까지 모든 방법을 동원해서 얼굴 애니메이션 작업을 하고 있다. 본 논문에서는 3차원 얼굴 애니메이션 분야에서 작은 부분이지만 얼굴 애니메이션 제작에 있어서 많은 부분에서 활용 가능성이 높은 새로운 방법을 제시하고자 한다.

현재 얼굴 애니메이션 제작에 있어서 가장 보편적으로 사용되어지고 있는 방법은 미리 모델링된 표정 기저 모델(Expression basis)들의 선형 조합으로 새로운 표정을 만들어 내는 블랜드쉐입 방법이라고 할 수 있다. Maya나 SoftImage와 같은 많은 상용 애니메이션 패키지에서 블랜드쉐입 기반 얼굴 애니메이션을 지원하고 있다. 본 논문에서 제안하고자 하는 기술도 이와 같은 시스템이다. 블랜드쉐입 기반 얼굴 애니메이션 시스템을 개발하는데 있어서 근본적인 문제는 어떻게 기저 표정 모델을 제어하는냐에 있다. 한가지 방법으로는 애니메이터로 하여금 직접 표정 기저들의 가중치 값을 조절하여 원하는 표정 시퀀스를 만들어 내도록 하는것이다. 또 다른 널리 사용되는 방법은 배우의 움직임 정보를 활용하여 얼굴 애니메이션이 되도록 하는 것이다. 이 방법으로 접근 할 경우, 기저 모델이 움직임 정보를 획득한 사람과 동일할 경우, 원칙적으로 본래 움직임은 재생산이 가능해진다. 애니메이션 제작시 이러한 재복원 작업이 불필요할 수 있지만, 개발자의 입장에선 블랜드쉐입 기술의 벤치마크로서 중요성을 지닌다. 원래의 움직임 정보를 정확히 재복원을 할 수 있다면, 다른 얼굴 애니메이션도 정확히 만들어 낼 수 있기 때문이다. 본 논문에서는 얼굴 애니메이션 시스템이 모션 데이터 기반 제어에 의해서 이뤄지며 동시에 그 결과를 필요시 수동으로 제어 혹은 수정하는 것을 전제로 하고 있다.

블랜드쉐입 기술을 개발하는데 있어서 해결해야 할 또 다른 근본적인 이슈는 어떻게 표정 기저 모델을 생성하느냐에 관한 것이다. 본 논문은 이 이슈와 관련이 되어있다. 많은 애니메이션 스튜디오에서 이뤄지는 보편적인 방법은 직관적으로 인식할 수 있는 수동으로 만들어진 표정 기저를 사용하는 것이다. 이 경우에 기저 모델로 사용되어지는 갯수는 가능한 모든 표정을 표현할 수 있도록 충분히 많은 갯수의 기저 모델을 가지고 있어야 한다. 수동으로 만든 기저 모델의 장점은 비교적 예측 가능한 결과를 만들어 낼 수 있다는 것이다. 하지만 선형적 조합으로 만들어 낼 수 있는 표정이 전체 표정 공간(Expression space)에서 일부분만을 표현할 수 밖에 없다는 단점을 지닌다. 이러한 점은 실제 얼굴 모션 데이터를 재생산해 내는 경우에 재복원 오차를 크게 만든다는 것

을 통해 표정 공간의 커버 영역이 낮다는 것을 확인할 수 있다.

블렌드쉐입 기반 움직임 데이터 복원의 관점에서 볼때, 기저 모델 생성하는 방법으로 이미 알려진 좋은 방법이 있다. 모션 데이터의 주성분 분석을 통해서 표정공간을 스팬하는 상호 직교하는 기저 모델을 자동으로 생성하는 방법이 그것이다. 이렇게 생성된 기저 모델을 기반으로 원래 모션의 재복원을 취하면 매우 정확한 결과를 만들어 낼 수 있다. 그러나 이렇게 생성된 기저 모델은 시각적으로 직관적이지 않아서 애니메이터들이 특정 표정을 생성하기 위해 선형적 조합을 하는데 어려움이 따른다.

따라서, 본 논문에서는 기저 모델의 표현 공간 커버 영역은 통계적 방법에 의해 생성된 기저 모델의 그것과 유사하면서도, 동시에 의미 있는 표정을 갖는 기저 모델을 생성하는 새로운 방법을 제안한다. 이 방법은 수동으로 생성된 기저 표정 모델들이 변형이 되더라도 시각적으로 원래의 모델과 유사하면서도 표정 공간의 커버 영역을 확장할 수 있다는 고찰에 근거를 두고 있다. 또한 생성된 기저 모델들은 반드시 상호 직교하지 않아도 여전히 공간을 스팬 할 수 있다는 확장된 조건에도 기반을 두고 있다.

본 논문의 구성은 다음과 같다. 2장에서는 얼굴 애니메이션과 관련된 기존 연구들에 대해서 언급하고, 3장과 4장에서는 각각 본 논문에서 해결하고자 하는 문제와 유사 고유 얼굴 모델을 생성하는 방법에 대해서 설명한다. 그리고 5장에서는 실험 결과에 대해서 설명하며, 마지막으로 결론을 맺는다.

2. Related works

얼굴 표정을 생성하는 방법에 대해서 Parke[18]의 개척적인 연구가 진행된 이후, 많은 연구들이 진행되어져 왔다. 얼굴 표정은 턱, 근육 및 피부와 같은 많은 요소들의 조합 결과로 보여진다. Terzopoulos 등을 비롯해 많은 연구자들이 얼굴 표정을 합성하기 위해서 물리 기반 기술을 개발하였다[22, 23, 15, 6, 21]. 본 논문은 이것과는 다르게 접근을 한다. 즉, 여러 개의 기저 표정 모델의 선형적 조합으로 새로운 표정을 합성하는 방법이다. 이처럼 얼굴 요소들의 물리적 관점이 아닌 보다 자연스런 결과를 얻기 위해 실제 얼굴 움직임 데이터를 활용하는 것이다. 본 장에서는 기존의 블렌드쉐입 기술과 모션 데이터 기반 얼굴 애니메이션 기술들에 대해서 간략히 언급하고자 한다.

블렌드쉐입 기술은 얼굴 표정을 생성하는데 널리 사용되어지고 있는 방법이다. Kouadio는 블렌드쉐입 기술을 실시간으로 얼굴 표정을 생성하기 위해서 적용했는데, 라이브 모션 데이터를 획득해서 이를 기반으로 각 표정 기저 모델들의 가중치값을 결정하도록 하였다[14]. Phigin은 사람 얼굴의 2차원 사진으로부터 사실적인 3차원 얼굴 표정 모델을 생성하였으며, 블렌드쉐입 기술을 이용해서 생성된 얼굴 표정들 사이의 부드러운 전이가 가능토록 하였다[19]. Blanz와 Vetter는 3차원 얼굴 모델 데이터베이스에서 선형적인 조합을 통해서 2차원 사진이 입력으로 주어졌을때, 3차원 얼굴 모델을 생성하는 방법을 제시하였다[1]. Choe와 Ko는 애니메이터로 하여금 얼굴의 독립적인 근육을 수축시켰을때에 해당하는 얼굴 표정을 수동으로 생성토록하여 그것들의 선형적 조합으로 새로운 얼굴 표정을 생성하도록 하였다[7].

블렌드쉐입을 기반으로 새로운 표정을 합성한 결과의 수준을 결정하는 중요한 요소는 사용되어진 기저 모델의 표정 공간을 커버하는 영역이 얼마나 되는지에 의존한다. 따라서 Chuang은 커버하는 공간을 확대하기 위해서 기저 모델을 주성분 분석법을 통해서 구하였다. 하지만 생성된 기저모델은 의미있는 얼굴 표정으로 인식되지 못한 단점을 지녔다[8]. Chao는 독립 성분 분석법을 통해서 새로운 기저 모델을 생성하고자 하는 시도를 하였다. 이를 통해 주성분 분석법이 결과보다는 의미있는 표정들로 구분이 되는 기저 모델을 생성할 수 있었지만, 여전히 사람들에게 친숙하면서 분명한 표정들로 구성되진 못했다. 그 결과로 이렇게 생성된 기저 모델을 기반으로는 고전적인 키프레임 제어가 쉽게 되진 않는 단점을 여전히 지니고 있다[5]. 그리고 Joshi는 얼굴의 특정한 부위별 변형을 위해서 블렌드쉐입을 적용했는데, 이를 위해서 얼굴의 키 표정들을 의미있는 블렌드 영역으로 자동적으로 분할하는 방법을 제안하였다[13].

William은 사람의 표정을 합성하기 위해서 움직임 데이터를 활용하는 방법을 소개하였다[26]. 이 방법은 고차원의 자유도를 가지는 얼굴 움직임을 제어하는데 사람의 실제 움직임을 활용함으로써 매우 효율적이라는 걸 보여줬다. 얼굴의 모션을 블렌드쉐입 기반으로 재복원을 시도했던 방법은 이미 앞서 언급되었다. Noh를 비롯해서 많은 연구자들이 모션 데이터를 새로운 캐릭터에 대해서 표정을 합성하기 위한 리타겟팅 관련 연구를 시도하였다[17, 20, 16]. 최근에 Vlasic은 다선형 모델(multilinear model)을 활용하여 얼굴의 표정 및 음성 애니메이션을 다른 캐릭터에 전이하는 시도를 하였다[24]. 그리고 모션 데이터 기반 얼굴 애니메이션의 또다른 형태로 음성 기반 애니메이션이 있다. Bregler와 Ezzat 등과 같은 많은 연구자들 이 이 분야에 대표적인 연구를 수행해 왔다[2, 10].

3. Problem Description

$\mathbf{v} = \mathbf{v}(t) = [\mathbf{v}_1^T, \ldots, \mathbf{v}_N^T]^T$를 시간 t에서의 3차원 얼굴 모델의 동적인 모양을 표현한다고 정의할 때, 이것은 N개의 버텍스로 구성된 삼각형 메쉬이다. 여기서 $\mathbf{v}_i$는 i번째 버텍스의 3차원 위치를 나타낸다. 또한 $\mathbf{v}^0 = [(\mathbf{v}_1^0)^T, \ldots, (\mathbf{v}_N^0)^T]^T$를 기준이 되는 얼굴 표정 모델로 표현하며, 얼굴 모션 데이터를 $\Xi = [\mathbf{v}(1), \ldots, \mathbf{v}(L)]$와 같은 $3N \times L$ 행렬로 표현한다. 여기서 L은 획득된 모션 데이터의 전체 프레임수이다. 본 논문에서 관심이 있는 것은 Ξ를 스팬하는 선형적 조합을 이루는 표정 기저들의 집합을 찾는 것이다.

애니메이터가 수동으로 생성한 표정 기저 모델 집합을 $\hat{E}^H = \{\hat{\mathbf{e}}_1^H, \ldots, \hat{\mathbf{e}}_n^H\}$ 라고 정의할 때, n 은 기저 모델 요소들의 수이고, $\hat{\mathbf{e}}_i^H$ 는 i번째 기저 모델의 기하정보이다.

$\mathbf{e}_i^H$를 기준 모델(neutral face)로부터의 변위량이라고 하면 $\mathbf{e}_i^H = \hat{\mathbf{e}}_i^H - \mathbf{v}^0$이다.

본 논문에서는 $E^H = \{\mathbf{e}_1^H, \ldots, \mathbf{e}_n^H\}$와 같은 변위량의 집합을 특별한 혼동이 없다면, 수동으로 생성한 표정 기저 모델로 기술하도록 한다. 가중치 w_i^H가 주어졌을 때, 새롭게 합성

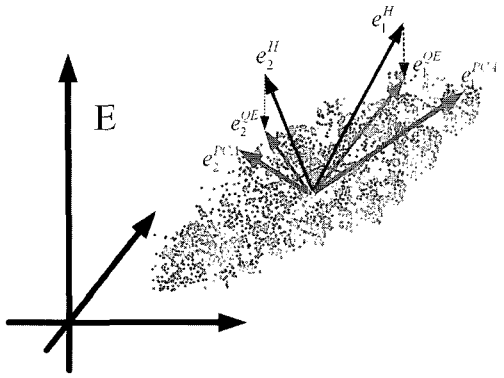

그림 1: Analogical drawing of the motion capture data and the bases discussed in the paper (red arrows: PCA basis, blue arrows: hand-generated basis, purple arrows: quasi-eigen basis)

된 얼굴 표정 $\mathbf{v}$를 아래와 같이 표현할 수 있다.

$$\mathbf{v} = \mathbf{v}^0 + \sum_{i=1}^{n} w_i^H \mathbf{e}_i^H. \quad (1)$$

여기서 수동으로 생성한 표정 기저를 사용했을 때 발생할 수 있는 문제는 그 기저 요소들의 선형적 조합이 Ξ을 커버할 수 있는 범위가 제한적이라는 것이다. 본 논문의 목적은 E^H를 새로운 기저 모델인 $E^{QE} = \{\mathbf{e}_1^{QE}, \ldots, \mathbf{e}_n^{QE}\}$로 변형함으로써, 이 새로운 기저 모델이 Ξ을 스팬하도록 하며 동시에 E^H의 해당하는 요소 모델인 $\mathbf{e}_i^H$와 시각적으로 유사하도록 하는것이다. 본 논문에서는 이러한 새로운 기저 모델을 유사-고유 얼굴 모델(*quasi-eigen faces*)이라고 한다.

4. Obtaining Quasi-Eigen Faces

만약 얼굴 모델의 버텍스들, $\mathbf{v}_1, \ldots, \mathbf{v}_N$가 3차원 공간상을 자유롭게 움직이게 한다면, $\mathbf{v}$는 $3N$-차원의 벡터 공간을 형성할 것이다. 본 논문에서는 이 공간을 *mathematical expression space* $\mathbf{E}$라고 일컫는다. 하지만 일반적인 얼굴 표정의 변화는 앞서 언급한 공간의 좁은 영역에 국한되어 있다. 만약 Ξ상에 존재하는 얼굴 표정 각각을 $3N$-차원 공간에 한 점으로 표현을 한다면, Point cloud는 근사적으로 hyperplane을 형성할 것이다. 주성분 분석법(principal component analysis)을 통하여, 형성된 초평면을 스팬하는 직교하는 좌표축을 구할 수 있게 된다. 이러한 상황을 그림 1에서 3차원으로 축소된 형태로 나타내었다. 3차원 좌표 시스템은 $\mathbf{E}$를 나타내며, 점들은 Ξ-hyperplane을 형성하고 있으며, 직교하는 굵은 좌표축은 주성분을 나타낸다.

유사-고유 얼굴 모델을 생성하는 과정은 주성분 분석에 기초를 두고 있다. 전체 모션 캡쳐 데이터 Ξ를 모두 더했을 때, $\mu = [\mu_1^T, \ldots, \mu_N^T]^T$는 $\mathbf{v}$의 평균값이 된다. 따라서, 중심 이동이 된 point cloud $\tilde{\mathbf{D}} = [\tilde{\mathbf{v}}(1)^T, \ldots, \tilde{\mathbf{v}}(L)^T]^T$를 얻게 된다. 여기서, $\tilde{\mathbf{v}}(i) = \mathbf{v}(i) - \mu$이다. 이제, 분산 행렬인 $\mathbf{C}$를 아래와 같이 구하게 된다.

$$\mathbf{C} = \frac{1}{L}\tilde{\mathbf{D}}\tilde{\mathbf{D}}^T. \quad (2)$$

$\mathbf{C}$는 대칭인 양정치 행렬이므로, 양의 고유값을 가지게 된다. $\mathbf{C}$의 고유값인 $\lambda_1, \ldots, \lambda_{3N}$를 크기 순으로 정렬을 해서, λ_1을 제일 큰 값이 되도록 한다. $\sum_{i=1}^{m}\lambda_i / \sum_{i=1}^{3N}\lambda_i$ 식에 의해서 미리 정한 커버율에 맞는 주성분의 갯수로 $\{\lambda_1, \ldots, \lambda_m\}$를 정하면, 이에 해당하는 m개의 고유 벡터인 $E^{PCA} = \{\mathbf{e}_1^{PCA}, \ldots, \mathbf{e}_m^{PCA}\}$가 구해진다. 이렇게 구해진 E^{PCA}에서의 표정 기저 모델을 고유 얼굴 모델(*eigenfaces*)이라고 부른다. 일반적으로 아주 작은 수의 m이라도 커버영역이 거의 100%에 가깝게 나오기 때문에 앞의 과정은 주어진 표정 집합인 Ξ를 커버하는 표정 기저를 생성하는 아주 좋은 수단이 된다. 참고로 본 논문에서는 m의 수를 18로 하여 Ξ의 99.5%를 커버하도록 하였다. 하지만 이러한 접근 방법은 고유 얼굴 모델이 수학적인 중요성을 지니긴 하지만, 시각적인 인지가 가능한 얼굴 모델을 표현하고 있지 않다는 단점을 지닌다.

고유 얼굴 모델의 생성을 기반으로, 이제 애니메이터가 수동으로 생성한 표정 기저 모델을 유사-고유 얼굴모델로 변형하는 것에 대해서 기술하고자 한다. 이 방법은 수동으로 생성된 표정 모델들이 hyperplane에서 벗어난 위치에 존재한다는 가정에 근거를 두고 있다. 그림 1에서 이를 유추할 수 있는 상황에 대해서 나타내었듯이, 두 개의 수동으로 생성된 표정 모델(그림에서 두 개의 3차원 벡터)이 hyperplane상에 존재하지 않는 것을 보여준다. 만일 이 두 개의 표정이 hyperplane상에 존재한다면, 원칙적으로 이 두 표정 모델의 선형 조합으로 모든 표정을 표현할 수 있겠지만, 그림에서처럼 hyperplane 밖에 존재하는 경우에 있어서는, 실질적으로 일차원의 범위만 스팬하게 되어 결과적으로 상당한 커버 영역의 손실을 만들게 된다. 이러한 제한 조건을 무시한 채 선형적 조합이 이뤄지면 그 결과는 hyperplane 바깥에 존재하게 된다. 이것이 수동으로 생성한 기저 모델을 사용하면 원모션 데이터의 재복원시 많은 오차를 포함하게 되는 이유이다.

앞서 언급한 문제를 푸는 간단한 방법은 수동으로 생성한 기저 모델들을 hyperplane 상으로 투영하는것이다. 즉, 본 논문에서 구하고자 하는 유사-고유 얼굴 모델은 수동으로 생성한 기저 모델 요소들을 투영하는 것이다. 수동으로 생성한 얼굴 기저 모델들을 각각의 주성분 축으로 투영하기 위해서 먼저 아래의 식에 따라 계산한다.

$$w_{ij}^{\text{PCA-to-QE}} = \mathbf{e}_j^{PCA} \cdot (\mathbf{e}_i^H - \mu), \quad (3)$$

여기서, i는 수동으로 생성한 기저 모델들의 범위를 나타내며, j는 전체 주성분 축의 범위를 나타낸다. 이제 유사-고유 얼굴 모델은 아래 식에 의해 구해진다.

$$\mathbf{e}_i^{QE} = \mu + \sum_{j=1}^{m} w_{ij}^{\text{PCA-to-QE}} \mathbf{e}_j^{PCA}. \quad (4)$$

유사-고유 기저 모델을 사용하여, 새로운 얼굴 표정은 $\sum_{i=1}^{n} w_i^{QE}\mathbf{e}_i^{QE}$과 같은 선형 조합으로 생성된다. 식 3과 4를 거친 투영 과정은 원래 수동으로 생성한 기저 모델 요소들을 변형시킨다. 본 논문에서는 새롭게 생성된 얼굴 기저 모델들이 시각적으로 원래의 것과 유사한지를 평가할 필요가 있다.

만일 수동으로 생성한 얼굴 표정 모델이 hyperplane상에 존재한다면(혹은, 모션 캡쳐 데이터에 포함이 되어 있다면), 그 모델은 투영 작업을 통해서도 아무런 변형이 생기지 않을것이다. 만일 수동으로 생성한 표정 모델이 hyperplane 바깥에 존재한다면, 투영 단계를 거치게 되면 최소의 유클리디언 변형이 생길 것이다. 시각적 차이가 유클리디언 거리와 동일한 스케일이 아니지만, 작은 유클리디언 거리는 일반적으로 작은 시각적 변화에 대응된다고 할 수 있다.

확인해야 할 또 다른 면은 E^{QE}의 커버하는 영역에 관한 것이다. 위의 그림 1에서, 만일 두개의 3차원 벡터가 일치하지 않으면 그 벡터의 투영된 결과로 생성된 벡터들은 hyperplane을 스팬할 가능성은 매우 높다. 유사한 예로, 수동으로 생성한 표정 기저의 수가 m보다 같거나 클 경우에는 그것들의 투영된 결과 역시 hyperplane Ξ를 커버 할 가능성은 매우 높다고 할 수 있다. 아래에는 실제로는 일어날 가능성이 희박하지만, 피해야 할 상황과 그에 대한 해결책을 기술한다.

Preventive Treatments: 애니메이터에게 수동으로 표정 기저 모델을 생성할 때, 기저 모델들이 투영될 때 중복 되지 않도록 미리 가이드라인을 제시해 준다. 예를 들어 얼굴의 근육들 중에서 하나의 표정 근육을 최대한 수축시키고 나머지 근육들은 이완시킨 채로 두었을때의 표정을 각각의 근육에 따라 $\hat{\mathbf{e}}_i^H$로 표현하도록 할 수 있다. 이것은 본질적으로 두 개의 수동으로 생성한 표정 기저 모델들이 거의 동일한 모양을 가지도록 하는 가능성을 배제하는 것이다 [7]. 이러한 목적을 위해서 애니메이터에게 개별 근육들의 각각의 수축된 모습을 보여주는 문헌[11]을 참고토록 하는 것이 비중첩 표정 기저 모델을 생성하도록 하는 좋은 방법 중에 하나이다.

Post-Treatments: 앞서 언급한 사전 처리에도 불구하고, 유사-고유 얼굴 모델이 주성분 축을 채우지 못하는 경우가 발생 할 수 있다. 이러한 경우의 상황은 행렬 $\mathbf{W}^{\text{PCA-to-QE}} = (w_{ij}^{\text{PCA-to-QE}})$를 확인함으로써 발생 유무를 알 수 있다. 만약 $\sum_{i=1}^{n} |w_{ij}^{\text{PCA-to-QE}}|$이 특정 문턱값 ϵ보다 작으면, 유사-고유 기저 모델들 중에서 $\mathbf{e}_j^{PCA}$에 해당하는 축이 비었다고 결정할 수 있게 된다. 이러한 경우, 우리는 단순히 $\mathbf{e}_j^{PCA}$를 기저 모델로 추가하거나, 애니메이터로 하려금 해당 고유 성분을 최소로 변형을 가하도록 하여 투영시킨 후, 유사-고유 모델로 추가하는 방법이 있다.

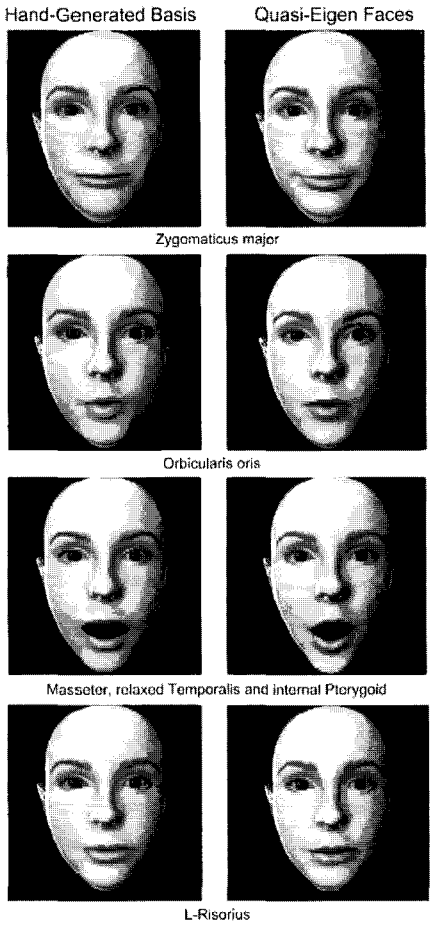

그림 2: Side-by-side comparison of the hand-generated basis and the quasi-eigen basis in four selected elements

5. Experiments

제안한 방법을 테스트하기 위해서, 본 논문에서는 얼굴 캡쳐 데이터를 획득하고, 표정 근육의 수축을 기반으로 수동으로 표정 기저 모델을 생성하였다. 앞 장에서 언급한 과정을 통해서 유사-고유 얼굴 모델을 수동을 생성한 모델로부터 구하였다.

5.1 Capturing the Facial Model and Performance

본 논문의 실험을 위해서, 바이콘 옵티컬 시스템(Vicon Optical System)을 이용하여 배우의 얼굴 표정의 움직임을 획득했다. 8대의 카메라를 이용하여, 초당 120Hz의 배우의 얼굴에 붙인 66개 마커의 움직임을 추적하였고, 추가로 7개의 마커를 배우의 머리에 부착하여 머리의 전체적인 움직임을 얻을 수 있도록 하였다. 캡쳐된 모션 데이터의 전체 길이는 L=35,000 이었다. 3차원 얼굴 모델을 생성하기 위해서 Cyberware 3차원 스캐너를 이용하였으며, 모션 캡쳐 시스템으로부터 획득한 3차원 마커의 위치와 스캐너를 통해서 생성된 3차원 기하 모델과의 대응점을 확립하기 위해서 Pighin이 소개한 방법을 사용하였다[19].

5.2 Preparing the Training Data Ξ

모션 캡쳐 데이터인 Ξ는 얼굴 기하 정보의 연속된 값이라고 할 수 있다. 우리는 각 프레임별로 획득한 마커의 위치 정보를 방사형 기저 함수(radial basis function)를 기반으로 한 보간법을 이용하여 얼굴 메쉬로 변형할 수 있다. 실제로 제안하는 방법은 만일 얼굴 모션 데이터가 메쉬 데이터로 주어지면 마커 데이터에 직접 적용도 할 수 있다. 이 두가지 접근 방

법 모두 같은 결과를 나타낸다.

5.3 Preparing the Hand-Generated Expression Basis

앞장에서 언급한대로 획득된 모션 데이터에 대해서 주성분 분석 처리를 해서 $m = 18$인 주성분 축을 취하여 Ξ의 99.5% 영역을 커버하도록 하였다. 우리는 애니메이터로 하여금 직접 표정 기저 모델을 18개로 하는 기저 집합 $\hat{E}^H = \{\hat{\mathbf{e}}_1^H, \ldots, \hat{\mathbf{e}}_n^H\}$를 만들도록 하였다. 만일 생성된 기저 모델들이 E상에서 어떤 영역에 모여 있을 경우에는 그것의 투영된 결과 역시 hyperplane상에 모이게 될 것이다. 이것은 결국 낮은 커버 범위를 만들게 되고, 결과적으로 추가적으로 수동으로 기저 모델을 생성하게끔 한다. 애니메이터의 수작업을 줄이기 위해서 우리는 표정에 관련된 근육의 위치와 크기를 고려하여 모델링하도록 가이드를 하고, 따라서 각 기저 모델이 하나의 표정 근육이 충분히 수축되고 나머지는 이완 되었을 때의 얼굴 모양과 대응되도록 하였다. 본 논문의 실험을 위해서 우리는 18개의 수동으로 생성한 얼굴 기저 모델을 사용하였고, 그 중에서 6개는 얼굴의 윗부분 근육의 움직임에 관련된 것이고, 나머지 12개는 얼굴의 아랫 부분 근육의 움직임에 관련된 모델이 되도록 하였다.

5.4 Obtaining the Quasi-Eigen Faces

수동으로 생성한 표정 기저로부터, 4장에서 설명된 단계를 따라서 유사-고유 얼굴 모델을 생성한다. 수동으로 생성된 얼굴 모델과 그에 대응하는 유사-고유 얼굴 모델로의 변형된 결과를 몇가지 표정에 대해서 그림 2에서 보였다. 수동으로 생성된 얼굴 표정 모델과 유사-고유 얼굴 모델의 비교를 통해서, hyperplane상으로 투영되어 적지 않은 기하 정보의 변형에도 불구하고 원래의 시각적 느낌은 그대로 유지되고 있음을 알 수 있다.

전처리 단계로써, 6000여 프레임의 학습 데이터에 대해서 주성분 분석 처리를 하는데 Intel 펜티엄 4 3.2GHz CPU, Nvidia geforce 6800 GPU 환경에서 158분이 소요 되었다. 데이터의 학습 과정이 끝나고 나면, 유사-고유 얼굴 모델은 실시간으로 생성된다.

5.5 Analysis

획득된 모션데이터의 재복원 과정을 통해 생성된 유사-고유 얼굴 모델의 커버 영역을 가늠할 수 있다. 이를 위해서 모션 데이터의 매 프레임마다 유사-고유 얼굴 모델의 선형적 조합을 구한다. $\mathbf{v} = \mathbf{v}^0 + \sum_{i=1}^{n} w_i^{QE}\mathbf{e}_i^{QE}$를 어떤 프레임에서의 재복원된 얼굴 모델이라고 하고, $\mathbf{v}^* = \mathbf{v}^0 + \mathbf{d}^*$를 Ξ에서의 원 표정 모델이라고 하자. 이때 $\mathbf{d}^*$는 중립 표정 모델로부터의 $3N$-차원의 변위 벡터이다. 이를 바탕으로 아래의 식 5을 최소화하는 n-차원 가중치 벡터인 $\mathbf{w}^{QE} = (w_1^{QE}, \ldots, w_n^{QE})$를 구할 수 있다.

$$|\mathbf{v}^* - \mathbf{v}|^2 = \sum_{j=1}^{N} \left| \mathbf{d}_j^* - \sum_{i=1}^{n} w_i^{QE}\mathbf{e}_{ij}^{QE} \right|^2, \qquad (5)$$

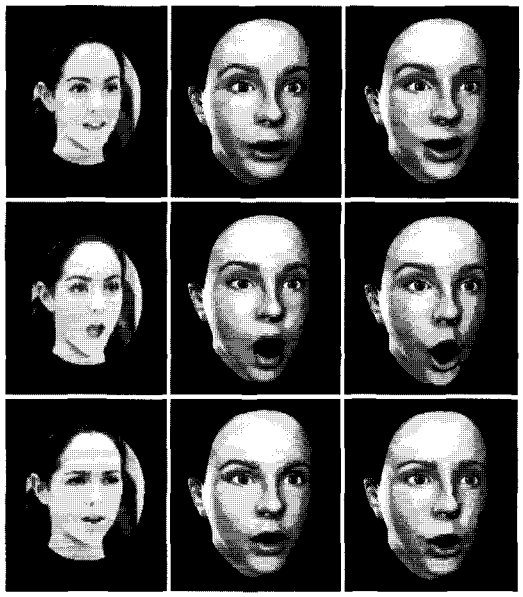

그림 3: Reconstruction of the original performance (left: motion capture, middle: with E^H, right: with E^{QE})

여기서 $\mathbf{d}_j^*$과 $\mathbf{e}_{ij}^{QE}$는 각각 $\mathbf{v}^0$로부터 버텍스 $\mathbf{v}^*$의 j-번째 변위값과 $\mathbf{e}_i^{QE}$의 변위값을 나타낸다. 식 5를 이차 계획법(quadratic programming)을 이용하여 푸는데 프레임당 0.007초가 소요되었다. 재복원의 정확도를 평가하기 위해서 다음과 같은 오차 측정식을 사용하였다.

$$\alpha[\%] = 100 \times \frac{\sqrt{\sum_{j=1}^{N} |\mathbf{v}_j^* - \mathbf{v}_j|^2}}{\sqrt{\sum_{j=1}^{N} |\mathbf{v}_j^*|^2}} \qquad (6)$$

비교를 위해, 상기 분석 과정을 E^H 와 E^{PCA}에 대해서도 수행하였다. 각각의 기저를 사용했을 때의 α 값은 각각 $\alpha^{QE} = 0.72\%$, $\alpha^H = 5.2\%$, 와 $\alpha^{PCA} = 0.62\%$ 이었다. 이 결과는 커버하는 범위의 관점에서 E^{QE}가 E^{PCA} 보다는 미세하게나마 모자라지만, E^H 보다는 월등히 낫다는 것을 보여준다.

그림 3는 재복원의 질적인 비교를 보여준다. 세로행 중에서 왼쪽 줄의 세 개의 영상은 캡쳐된 원래의 움직임이며, 가운데와 오른쪽 줄의 세 개의 영상은 각각 E^H 와 E^{QE}의 기저 모델을 사용해서 만든 재복원 영상이다. 그림 4은 그림 3의 제일 위쪽에서 보여준 프레임에 대해서 재복원을 했을때 생기는 오차를 시각화한 것이다. 여기서 빨간색 점은 캡쳐된 마커의 위치를 나타내고, 파란색 점은 재복원 되었을때의 결과 위치를 나타낸다. 여기서, 오른쪽 그림에서 보이는 유사-고유 얼굴 모델을 기반으로 재복원한 것이 수동으로 만든 기저 모델을 사용한 것보다 훨씬 오차가 적다는 것을 확인 할 수 있다. 보다 동적인 얼굴 움직임에 대한 재복원의 질적 비교를 위해서 본 논문에서는 배우가 말하는 장면을 재복원하

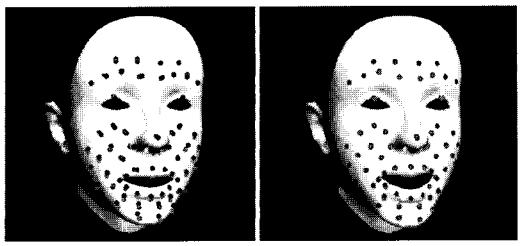

그림 4: Comparison of the reconstruction errors (left: for E^H, right: for E^{QE})

는 실험을 하였다. 이 결과 제안하는 방법으로 만든 기저 모델을 사용하는 것이 수동으로 만든 기저 모델을 기반으로 하는 것보다 훨씬 더 자연스러운 결과를 낼 수 있었다.

6. Conclusion

본 논문에서는 블랜드쉐입 기반 얼굴 애니메이션 시스템에 대해서 표정 기저 모델들을 생성하는 새로운 방법을 제시한다. 애니메이션 스튜디오에선 보편적으로 그러한 기저 모델들을 수동으로 만들어서 사용하는데, 수동으로 생성한 표정 기저는 실제 얼굴 표정과는 다른 요소를 가질 수 있기 때문에 이들을 기반으로 선형 조합을 취해서 재복원 및 애니메이션을 만들면 재복원시 오차와 비사실적인 결과를 만들 수 있다. 반면, 통계적 접근 방법으로 생성한 기저 모델들은 매우 사실적인 결과를 만들어 낼 수 있지만, 생성된 기저 모델들이 직관적인 인지가 어렵다는 단점을 지닌다. 따라서 본 논문에서는 유사-고유 얼굴 모델을 생성함로써, 직관적 인지가 가능하여 애니메이터로 하여금 수동 제어가 가능하도록 하며 동시에 수동으로 생성한 기저 모델과 비교하여 모션의 재복원 오차를 현저히 낮출 수 있었다.

본 논문에서는 획득된 모션의 재복원에 중점을 두었다. 얼굴 애니메이션에서의 실제 경험에 기반을 이 접근 방법은, 기술적인 문제점이 대개 합성 부분에서라기 보다는 분석처리하는 부분에서 문제점을 가지고 있다는 사실에 초점을 두었다. 만일 분석처리가 정확히 이루어진다면, 표정의 합성은 그것이 재복원이든 다른 캐릭터에서의 애니메이션이든 정확히 이루어지게 될 것이다. 본 논문에서 이뤄진 실험은 제안한 방법이 일반적으로 인지할 수 있는 사람의 표정을 가진 기저 모델을 생성하고 이와 더불어 분석처리 된 후 재복원 오차가 현저히 줄어들었다는 것을 보여준다.

제안하는 방법은 그 결과가 애니메이터에 의해 제공되어지는 수동으로 생성한 표정 기저에 의존하는 animator-in-the-loop 방법이다. 만일 애니메이터가 불충분한 표정 기저를 제공한다면, 그것을 기반으로 해서 새롭게 생성된 표정 기저 역시 더 나은 결과를 제공치 못할 것이다. 본 연구에서는 수동으로 표정 기저 모델을 생성하는 방법이 5장에서 언급되었듯이 근육 기반으로 접근하는 것이 기저 모델의 커버 영역을 확장시키는데 효율적이라는 것을 확인하였다. 하지만 이 접근 방법이 커버 영역의 중첩을 막으면서, 수동으로 표정기저를 생성하는 유일한 방법이 아니므로, 이 부분에 대해서 향후에는 더 나은 방법이 고려되어져야 할 것이다.

Acknowledgment

이 연구는 과학기술부 국가지정연구실사업(M10600000232-06J0000-23210), 정보통신부 선도기반기술사업, BK21, 서울대학교 자동화시스템공동연구소의 지원으로 수행되었음.

참고 문헌

[1] BLANZ, V. AND VETTER, T. 1999. A morphable model for the synthesis of 3D faces. In *Proceedings of SIGGRAPH 1999*, ACM Press, 187–194.

[2] BREGLER, C. 1997. Video Rewrite:driving visual speech with audio. In *Proceedings of SIGGRAPH 1997*, ACM Press, 353–360.

[3] CHAI, J., XIAO, J. AND HODGINS,J. 2003. Vision-based control of 3D facial animation. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 193–206.

[4] CHANG, Y.-J., AND EZZAT, T. 2005. Transferable Videorealistic Speech Animation. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 143–151.

[5] CHAO, Y., FALOUTSOS, P. AND PIGHIN, F. 2003. Unsupervised learning for speech motion editing. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 225–231.

[6] CHOE, B., LEE, H AND KO, H.-S. 2001. Performance-driven musclebased facial animation. In *Journal of Visualization and Computer Animation 1999*, 67–79.

[7] CHOE, B. AND KO, H.-S. 2001. Analysis and synthesis of facial expressions with hand-generated muscle actuation basis. In *Proceedings of Computer Animation 2001*, 12–19.

[8] CHUANG, E. 2002. *Performance driven facial animation using blendshape*. Stanford University Computer Science Technical Report, CS-TR-2002-02, 2002.

[9] EKMAN, P. AND FRIESEN, W. V. 1978. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Inc.

[10] EZZAT, T., GEIGER, G., AND POGGIO, T. 2002. Trainable videorealistic speech animation. In *Proceedings of SIGGRAPH 2002*, ACM Press, 388–398.

[11] FAIGIN G. 1990. The Artist's Complete Guide to Facial Expression. Watson-Guptill Publications

[12] GUENTER, B., GRIMM, C., WOOD, D. MALVAR, H. AND PIGHIN, F. 1998. Making faces. In *Proceedings of SIGGRAPH 1998*, ACM Press, 55–66.

[13] JOSHI, P., TIEN, W., C., DESBRUN, M., AND PIGHIN, F. 2003. Learning controls for blend shape based realistic facial animation. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*

[14] KOUADIO, C., POULIN, P., AND LACHAPELLE, P. 1998. Real-time facial animation based upon a bank of 3D facial expression. In *Proceedings of Computer Animation 1998*, IEEE Computer Society Press.

[15] LEE, Y., TERZOPOULOS, D., AND WATERS, K. 1995. Realistic modeling for facial animation. In *Proceedings of SIGGRAPH 1995*, ACM Press, 55–62.

[16] NA, K. AND JUNG, M. 2004. Hierarchical retargetting of fine facial motions. In *Proceedings of Eurographics 2004*, Vol. 23.

[17] NOH, J. AND NEUMANN, U. 2001. Expression cloning In *Proceedings of SIGGRAPH 2001*, ACM Press, 277–288.

[18] PARKE, F. I. 1972. Synthesizing realistic facial expressions from photographs. In *Proceedings of ACM Conference 1972*, ACM Press, 451–457.

[19] PIGHIN, F., HECKER, J., LISCHINSKI, D., SZELISKI, R., AND SALESIN, D. H. 1998. Synthesizing realistic facial expressions from photographs. In *Proceedings of SIGGRAPH 1998*, ACM Press, 75–84.

[20] PYUN, H., KIM, Y., CHAE, W., KANG, H. W., AND SHIN, S. Y. 2003. An example-based approach for facial expression cloning. In *Proceedings of ACM SIGGRAPH/Eurographics Symposium on Computer Animation.*, 167–176.

[21] SIFAKIS, E., NEVEROV, I., AND FEDKIW, R. 2005. Automatic determination of facial muscle activations from sparse motion capture marker data. In *Proceedings of SIGGRAPH 2005*, ACM Press, 417–425.

[22] TERZOPOULOS, D., AND WATERS, K. 1990. Physically-based facial modeling, analysis and animation. *The Journal of Visualization and Computer Animation*, 73–80.

[23] TERZOPOULOS, D., AND WATERS, K. 1993. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 6, 569–579.

[24] VLASIC, D., BRAND, M., PFISTER, H., AND POPOVIC, J. 2005. Face transfer with multilinear models. In *Proceedings of SIGGRAPH 2005*, ACM Press, 426–433.

[25] WATERS, K. 1987. A muscle model for animating three-dimensional facial expressions. In *Proceedings of SIGGRAPH 1987*, ACM Press, 17–24.

[26] WILLIAMS, L. 1990. Performance-driven facial animation. In *Proceedings of SIGGRAPH 1990*, ACM Press, 235–242.