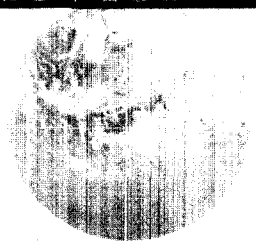
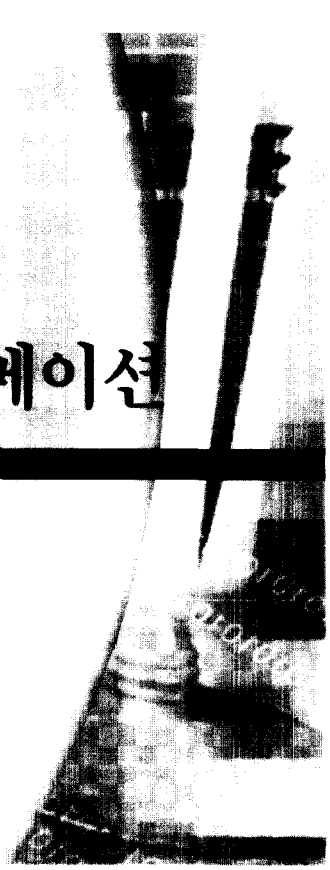
3차원 영상 합성과 그래픽 애니메이션

□ 전 성 희 · 임 동 곤 · 홍 유 성 : 광주과학기술원 정보통신공학과





1. 3차원 영상 합성

1.1 3차원 영상 합성의 목적

최근 지식기반 사회의 주연으로 등장한 영상 산업은 국가 경제는 물론이고 문화의 급속한 변화를 촉진시키고 있다. 영상 산업은 과학, 기술, 군사, 언론, 문화, 의료 등 여러 분야에서 인간 생활에 밀접한 관련이 있다. 그러나 현재 주로 사용하고 있는 2차원 영상으로는 인간의 다양한 욕구를 충족시키기에는 한계가 있기 때문에 3차원 영상의 이용은 필연적이며, 최근 이에 대한 관심이 급격히 증가하고 있다. 이와 같은 3차원 영상에 대한 표현 방법은 흔히 컴퓨터 그래픽 기술을 이용해 객체를 표현하거나 입체사진(Stereoscopic Image)을 통해 실감나게 표현하기도 하며, 3차원 실사 영상(Real Image)과 3차원 그래픽 영상을 자연스럽게 합성하여 실시간으로 생생하게 합성 영상을 제공하려는 다양한 방법들이 시도되고 있다. 즉, 카메라로부터 저장된 배경 영상에 임의의 객체(Object)를 합성하거나 컴퓨터 그래픽으로 만든 배경에 실사영상을 합성하여 몰입감(Immersion)이나 현실감(Reality)을 높이려는 것이 주요한 목적 중에 하나이다.

1.2 영상 합성 분야

영상 합성 기술은 크게 컴퓨터 그래픽 분야에서의 접근과 컴퓨터 비젼 쪽에서의 접근 방법이 있으나, 주로 스테레오 비젼의 정합에 근거한 접근이 많이 연구되고 있다. 스테레오 비젼에서 실재감과 현실감을 높여주는 다시점 영상표시 방법은 스테레오 영상으로부터 대응점을 찾아 삼각측량 방법으로 카메라에서 물체까지의 거리 정보를

구한 다음, 거리 정보를 삼각측량법에 다시 적용하여 비디오나 카메라로 입력된 영상들 중에 하나인 중간 위치에 있는 영상에 대한 대응점의 위치 좌표를 계산하는 것이다. 그러나 일반적인 거리 정보의 추출과는 달리 중간 영상의 모든 화소에 대한 거리 정보를 구해야 하므로 정확하고 조밀한 거리정보 추출이 필요하다. 또한 자연스러운 실시간 동영상을 지향하는 경우에는 처리 속도의 비중도 매우 중요하다. 더불어 영상 획득을 위한 카메라의 배치 문제에 있어서도 자유로운 카메라의 배치를 갖는 영상 합성 알고리즘이 필요하다.

컴퓨터 그래픽의 입장에서 영상 합성은 그래픽이 놓일 위치를 결정하는 문제, 그래픽 가상 카메라를 실제 영상의 카메라 시각과 일치시키는 문제, 그래픽과 실제영상내의 물체간 상호 작용, 즉 그래픽 물체가 실물에 가려지거나 충돌할 경우에 관한 문제들을 해결해야 하는데 이와 같은 내용을 컴퓨터 그래픽 입장에서 주로 연구하는 분야가 증강현실(Augment Reality)이다[4].

실시간 영상 합성 기술은 3차원 TV 분야에서 절실히 요구되는 기술로서 고도의 영상처리 기술과 카메라 트래킹, 영상 변환 및 영상 합성 기술이 필요하다. 일본에서는 1990년대 초반부터 TAO(Telecommunication Advancement Organization)을 중심으로 TV 분야에 지속적인 투자와 연구가 이루어지고 있으며, 일본 NHK 연구소와 CRL 연구소에서는 2002년 월드컵 축구 경기를 TV로 방송하기 위해 준비하고 있으며, 일부 개발이 완료된 상태이다. 유럽에서는 ACT와 RACE 프로젝트의 일환으로 안경식 다안 3차원 전송시스템(DISTMA), 무안경식 다안 3차원 전송시스템(PANORAMA), 유선 원격 멀티미디어(MAESTRO), 3차원 TV 원격 수술 기술 개발(MIDSTER), 3차원 TV 영상물 제작 장비 기술 개발(MIRAGE) 등의 연구개발 과제가 수행되거나 완료되어 있다. 미국에서는 MIT, Stanford, CMU, Columbia 등의 대학과 JPL, Bell Labs 등의 연구소를 중심으로 입체영상 처리에 관련된 연구를 활발히 수행하고 있다. 국내에서도 이러한 3차원 TV 기술의 중요성을 인식하고 KIST를 비롯하여 몇몇 가전업체를 중심으로 3차원 TV 기술 개발을 착수해 일부 제품들을 개발하고 있다.

2. 스테레오 비젼과 영상 합성 기술

2.1 영상합성을 위한 스테레오 비젼

스테레오 비젼은 기본적으로 각기 다른 위치에 있는 한 대 이상의 카메라로부터 획득한 영상들을 분석하여 거리 정보를 구한다. 그러므로 거리 정보를 구하기 위해서는 한 영상에 존재하는 화소가 다른 영상의 어느 화소와 서로 대응되는지를 결정해야만 되는데, 이를 대응문제(Correspondence Problem) 또는 스테레오 정합 문제(Stereo Matching Problem)라고 한다. 이러한 정합 문제를 해결하기 위해서 수많은 연구가 진행되었으며 지금도 보다 정확한 정합을 결정하기 위해서 새로운 정합 알고리즘들이 제안되고 있다. 스테레오 정합에 대한 지금까지 연구 결과는 대부분 정합의 정확도를 높이기 위해서 몇 가지 제한 조건을 두어 스테레오 정합 알고리즘을 제시하였다. 즉, 에피폴라 제한 조건(Epipolar Constraint), 변이 경도 한계(Disparity Gradient Limits), 부드러운 표면 제한조건(Surface Smoothness Constraint) 등과 같은

제한 조건들은 스테레오 정합 알고리즘의 품질과 신뢰도에 강력한 영향을 미치는 요소들이다. 일반적으로 스테레오 영상에 대한 정합 방식은 크게 영역기반 방식과 특징기반 방식, 두 가지로 나누어진다. 영역기반 방식은 스테레오 영상의 영역을 정합하는 방식으로 모든 화소 마다 화소를 둘러싼 정합 창을 이용해 정합을 결정하는 반면에 특징기반 방식은 스테레오 영상에 있는 윤곽선(Edge)이나 모서리 등의 특징을 추출하여 특징간의 정합을 한 뒤에 전체 영상을 보간하는 방식이다. 영역기반 방식은 특징기반 방식에 비해 거리정보가 조밀한 반면에 계산량이 많다는 단점을 가지고 있다. 그러나 특징기반 방식은 계산량은 적지만 영상의 종류에 따라 민감하게 반응하고 특징이 아닌 부분에 대한 보간 방법이 완전하지 않기 때문에 영역기반 방식이 주로 이용되고 있다 [5]. 일반적으로 정합 점의 깊이 정보는 그림 1의 스테레오 기하를 통하여 (1)식을 간단하게 유도할 수 있다.

$$z = \frac{baseline \times f}{disparity} \qquad (1)$$

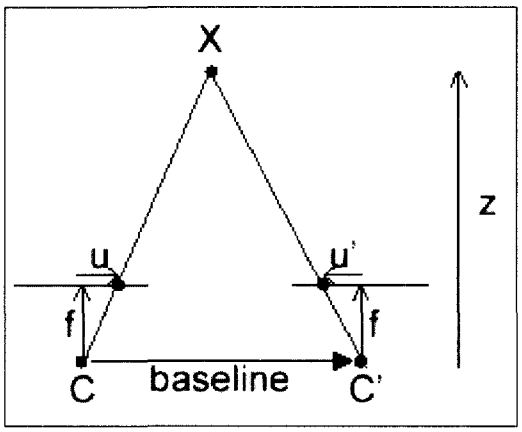


그림 1. 스테레오 기하학과 깊이 측정

(1)식에서 거리(z)는 렌즈 중심간의 거리인 베이스라인(Baseline)과 초점거리(f)를 변이로 나눈 값이다. 베이스라인과 초점거리는 알고있는 값이므로 변이(Disparity = u-u')만 결정하면 거리를 결정할 수 있다. 즉, 변이는 서로 대응하는 화소의 차이로서 거리를 결정하는데 중요한 요소가 된다. 일반적으로 스테레오 정합은 카메라 기하(Camera Geometry) 정보를 안다고 가정해 어느 정도 에러를 감수하고 정합을 구하거나 아니면 영상을 교정(Rectification)하여 거리 정보를 추출한다. 스테레오 정합 알고리즘을 통해 구한 변이를 카메라와 떨어진 거리에 따라 회색 영상(Gray Image)으로 표현한 것을 변이 지도(Disparity Map) 또는 깊이 지도(Depth Map)이라고 한다.

그림 2에 스테레오 영상에 대한 깊이 지도의 예가 있다. 깊이 지도는 카메라와 가까워 질수록 흰색에 가깝도록 표현하고 멀어질수록 검정색으로 표현한다. 스테레오 정합을 결정하기 위해서 적당한 크기의 창(Window)을 씌워 정합점을 결정하는데 그림 2(b)와 그림2(c)에 나타난 것과 같이 정합창의 크기는 정합의 신뢰도를 위해 충분히 광도(Intensity) 변화를 포함할 수 있을 크기가 되어야 하고 투영왜곡(Projective Distortion)를 피할 수 있을 만큼 또한 작아야 한다. 만일 정합창의 크기가 너무 작아서 광도 변화를 탐지하지 못하면 변이 추정의 신뢰도가 떨어진다. 그러나 정합창이 너무 커서 변이가 변하는 지역까지 씌우게 된다면 투영왜곡 때문에 역시 정확한 스테레오 정합을 결정하지 못한다. 영상의 광도 변화보다 큰 정합창을 스테레오 정합 알고리즘에 적용하게 되면 물체의 경계 부분과 같이 거리가 급격히 변하는 지역은 텍스쳐(Texture)

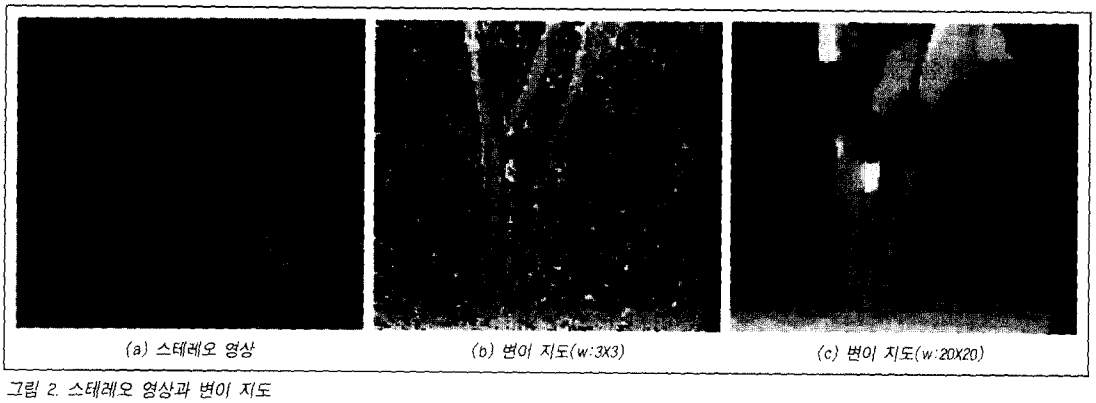


그림 2. 스테레오 영상과 변이 지도

가 많은 영역의 깊이 값이 적은 영역으로 스며들어 깊이 지도의 경계선 부분이 연장되는 현상(Fattening 또는 Boundary Overreach)이 발생된다[5]. 그렇기 때문에 정합창의 크기는 국부적인 광도와 변이에 의존해 적절하게 선택되어야만 한다. 이러한 문제점을 해결하기 위해서 Takeo Kanade와 Masatoshi Okutomi는 정합창의 크기를 변이의 신뢰도에 따라 유연하게 적용하는 방법을 제시하였다[1]. 또한 Daniel Scharstein과 Richard Szeliski는 비선형 분포 알고리즘을 제안하여 이런 문제점들을 개선하였다[2].

스테레오 정합의 문제점들은 주로 폐색(Occlusion), 깊이 불연속(Depth Discontinuity), 반복성(Regularity)으로부터 발생된다. 폐색은 실세계의 한 점 P가 좌측영상에 PL로 투영 되었을 때 우측 영상에서도 대응하는 점이 나타나야 하나 시야각 때문에 보이지 않을때 발생하는 문제이다. 이것은 주로 물체의 경계선부분에서 발생된다. 깊이 불연속은 대부분의 스테레오 정합 알고리즘이 전체 영상에 연속성(Continuity)을 사용함으로 발생되는 문제로서 영상의 깊이가 급격하게 변화되어 변이 추정에 어려움을 주는 요소이고 폐색과 깊은 관련이 있다. 반복성은 영상을 획득할 때 조명이나 주변 환경의 영향으로 인해 영상의 광도 변화가 규칙적으로 반복되거나 균일함으로 인해 정확한 정합을 결정하지 못하는 경우이다. 스테레오 비젼에서 정합을 결정하는데 어려움을 주는 위와 같은 문제점들은 지난 수많은 연구결과 이제는 어느 정도 만족할 만한 기술적인 수준까지 접근했다. 그러나 근래에 주요한 관심이 되고 있는 실시간 스테레오 정합으로 생성된 깊이 지도와 컴퓨터 그래픽 영상을 합성하는데는 위와 같은 제약들을 모두 극복하고 모든 화소에 대한 정확한 깊이를 할당했다고 가정했을 때 조차도 몇 가지 문제점이 여전히 남게 된다. 그 중에 하나가 영상 합성을 위해 깊이 지도로부터 배경과 전경(Foreground)을 분리했을 때 발생되는 후광(Halo)을 제거하기 위한 매트(Matte) 알고리즘이 필요하게 된 것이다. 즉, 스테레오 비젼이 그래픽 영상 합성과 같은 응용에 사용될 경우에 과거 스테레오 로봇 응용에서도 표현되지 않았던 몇 가지 요구 사항이 존재하게 된다. 우선 스테레오 정합 알고리즘은 깊이 불연

속인 지역을 포함해 모든 픽셀에 정확한 깊이를 할당할 수 있어야 한다. 그리고 배경과 전경을 각 픽셀의 진짜 색상을 정확하게 표현하면서 분리가 가능해야 하고 몇몇의 화소를 잃어버릴지라도 새로운 뷰(Novel View)을 만들고 정합 처리시 부분적으로 폐색된 지역들을 계산할 수 있어야 하는 등의 요구사항이 발생된 것이다. 이런 실사 영상과 그래픽 영상의 합성은 그래픽 분야에서도 현재 주요한 관심거리이다. 컴퓨터 그래픽적인 입장에서 볼 때 실사 영상과 그래픽 영상을 실시간으로 합성하기 위한 연구분야가 증강 현실이다. 증강 현실에서도 자연스러운 합성을 위해서 그래픽 영상이 실제 비디오 영상에 놓일 위치를 결정하는 문제나 그래픽과 실제 영상내의 물체간의 겹쳐질 경우 발생되는 색상 문제 등을 해결해야 하는 과제를 스테레오 비젼과 마찬가지로 안고있다. 이런 문제점들을 해결하기 위한 방안으로서 체적 재표현(Volume Representation), 계층적 표현(Layered Representation), 다중 깊이 지도(Multiple Depth Map)등의 방법을 이용하여 해결하려고 노력하고 있으며 Richard Szeliski가 여기에 대한 내용을 개략적으로 정리하였다[3].

스테레오 비젼에 대한 최근 연구결과에서 프랑스 INRIA의 Frederic은 1994년에 그림 3과 같이 사람의 얼굴을 스테레오 영상을 통해 3차원 모델링(Modeling)하였고 [6], 그림 4는 1999년에 미국 카네기 멜론 대학의 Takeo Kanade가 51대의 비디오 카메라가 부착된 돔형의 공간 안에서 움직이는 물체를 3차원 복원한 장면이다. 또한 그는 스테레오 정합을 통해 복원한 영상과 컴퓨터 그래픽 영상을 제트 키잉(Z-keying) 처리를 이용하여 실시간으로 합성 하였다[7]. 그림 5와 같이 카네기 멜론 대학에서 개발한 제트 키잉 처리를 이용한 방법은 실제 장면의 각 화소 정보, 즉 스테레오 정합으로 추출한 깊이정보를 이용하는 영상 키잉(Image Keying) 방법이다. 이때 제트 키는 실제 영상과 그래픽 영상의 깊이정보를 비교하여 카메라와 더 가까운 픽셀 값만을 표현하는 스위치 역할을 수행함으로써 각 화소에 대한

그림 3. 얼굴 모델링(Frederic Devernay, 1994 INRIA)

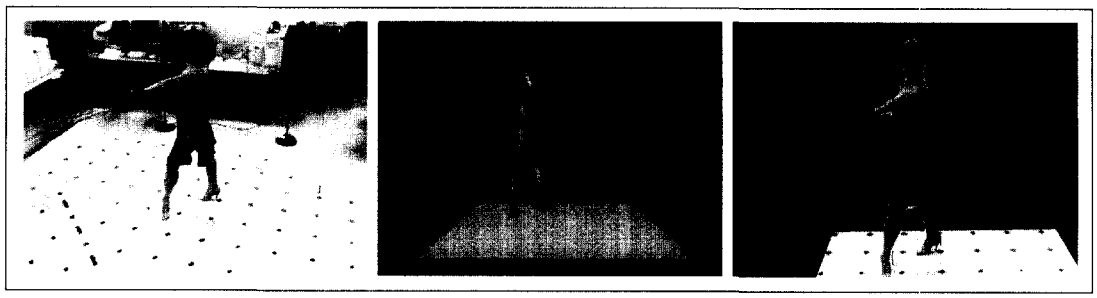


그림 4. 가상현실(Takeo Kanade, 1999 CMU)

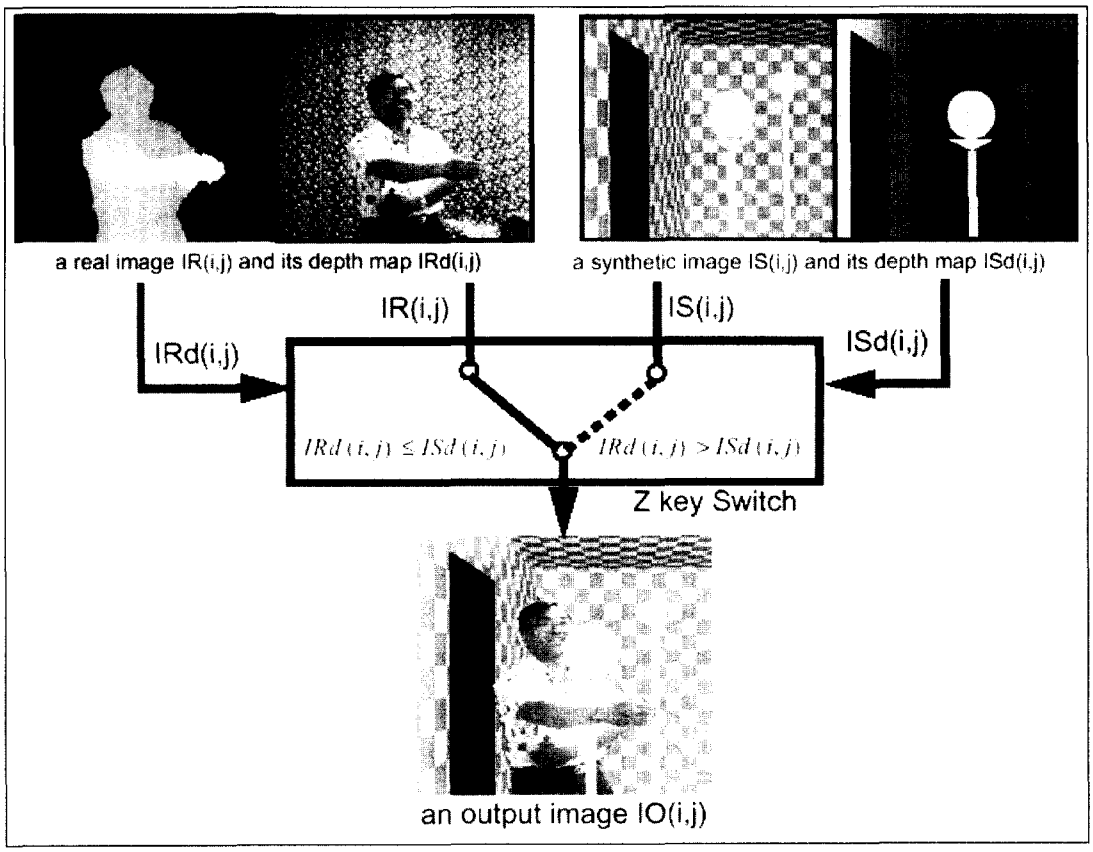


그림 5. Z-keying 처리를 이용한 영상합성(Takeo Kanade, 1995 CMU)

전경과 배경을 분리하여 새로운 시점의 영상(View)을 생성하는 것이다. 만일 실사 영상과 그래픽 영상간에 폐색이 발생하더라도, 즉 실제 영상의 물체가 그래픽 영상의 물체를 가리게 될지라도 가상 영상을 만들 수 있는 방법을 소개하였다. 그러나 최근 이와 같은 고무적인 현상에도 불구하고 실 시간 영상합성 방송을 위한 실질적인 제품으로 연결되기는 아직 많은 점을 보완 해야만 하는 실정이다. 현재의 알고리즘으로 만들어진 합성영상은 전경에 존재하는 물체 주변에 후광이 발생되며, 설사 스테레오 알고리즘으로 추정한 깊이정보가 정확하게 계산되었을지라도 전경과 배경 색상을 결합(Mixed Pixel)되는 픽셀들의 색상을 정교하게 제어하지 못하고 있다. 이런 현상은 특히 깊이 불연속인 지역에 있는 거의 모든 픽셀에서 발생된다. 그래서 제트 키잉이나 가상현실과 같은 응용 분야의 연구가 최근에 컴퓨터 그래픽 분야의 영상기반 랜더링(Image-based Rendering)과 관련해서 연구가 이루어지고 있다. 마이크로 소프트 연구소의 비젼 기술 그룹과 미국의 워싱톤 대학, 카네기 멜론대학, 유럽의 ECRC, 일본의 NHK 연구소와 CRL 연구소등에서 연구가 이루어지고 있다.

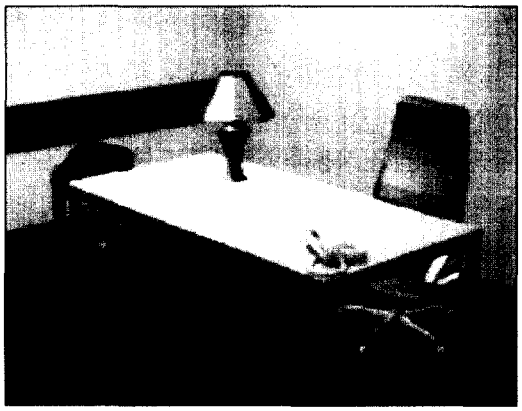


그림 6. 영상합성 (Courtesy David Breen, 1994 ECRC)

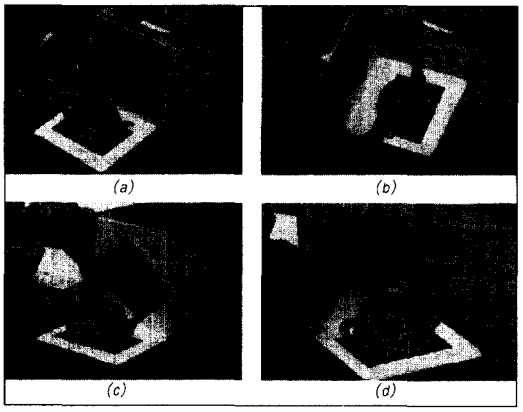


그림 7. 캘리브레이션을 하지 않는 영상합성(Kultulakos and Valino, 1998 Rochester University)

2.2 증강 현실과 TV 방송

증강현실 분야에서 1994년 ECRC(European Computer-Industry Research Center)의 Courtesy David Breen은 실사 영상에 두개의 가상 의자와 전등을 그림 6과 같이 합성했다. 미국 Rochester 대학의 Kultulakos와 Valino는 1998년에 카메라 캘리브레이션이나 물체에 대한 3차원 정보 없이 구현하는 방법을 제안하고 결과를 그림 7과 같이 보였다[8]. 제안한 방법은 초기화 과정에서 4개 이상의 기준점을 수작업으로 지정해야 되는 제한이 있으나 속도가 빠른 장점이 있다. 방송분야 실시간 영상합성은 가상 스크린(Virtual Screen) 시스템, 가상광고(Virtual Advertizing), 3차원 TV 방송 등이 스테레오 비젼과 증강현실 연구분야와 관련되어 현재 활발히 연구되고 있다.

3. 얼굴 애니메이션

3.1 얼굴 애니메이션의 역사

기본적인 얼굴의 애니메이션과 얼굴 모델에 대한 많은 연구가 진행되어 왔다. 최초의 모델은 Parke(1975, 1982)에 의해서 제안되었는데, 추가적인 매개변수를 가지고 구성되는 모델이다[9]. Platt와 Badler(1981)는 간략화 시킨 근육에 의해서 연결된 피부들의 연결구조에 힘을 가함에 의해서 표현되는 얼굴 감정을 표현한다. 그들의 시스템은 Ekman과 Friesen(1975)에 의해서 제안된 FACS(Facial Action Coding System)에 기반을 두고 있다[10]. Waters(1987)는 두 가지 형태의 근육을 가진 얼굴 모델을 개발했다[11,12]. 하나는 당김에 사용되는 선형근(Linear/Parallel)이고, 다른 하나는 조임에 사용되는 괄약근(Sphincter)이다.

Nahas(1988)는 B-스플라인에 기반한 방법을 제안하였다. 얼굴의 움직임은 제어점(Control Point)들의 움직임에 의해서 얻어진다. Magnenat-Thalmann(1988)은 추상적 근육 움직임(Abstract Muscle Action; AMA) 프로시져라고 하는 방법을 사용하여 근육 움직임을 구성하는 또 다른 접근법을 사용하였다[13].

Terzopolos와 Waters(1990)는 피부(Skin), 피하 지방 조직(Subcutaneous Fatty Tissue),

근육(Muscle)의 3가지 얼굴 조직들(facial tissues)을 사용하여 변형 가능한 3계층의 격자 구조를 가진 물리학에 기반 한 모델(Physics-based Model)을 제안하였다 [12]. Parke(1991)는 이전에 제안된 서로 다른 모델들의 서로 다른 매개변수를 사용한 구조를 재분석하고 이상적인 제어점 매개변수와 인터페이스 구조에 대한 미래의 방향을 제시하였다. Kalra(1991)는 다중 계층 구조를 사용하였는데, 여기서 각 레벨에서는 추상도(Degree of Abstraction)는 증가한다. 그는 또한 1992년에 비례적인 자유형 변형을 사용하여 얼굴 피부 표면을 변형시키는 접근법을 사용하였다. DiPaola(1991)는 얼굴 형의 범위의 확장을 허용하는 얼굴 애니메이션 시스템을 제안하였다.

또한 Lewis와 Parke(1987), Hill(1988), Magnenat-Thalmann(1987), Lewis(1991)등 여러 명의 저자들에 의해서 입 동기화(Lip Synchronization)와 음성 자동생성(Speech Automation)을 제공하는 연구가 이루어졌다.

최근의 여러 저자들은 인간의 모습으로부터의 정보를 이용한 새로운 얼굴 애니메이션 기술을 제안하였다. 추출된 정보는 얼굴 애니메이션을 조정하기 위하여 사용된다. 여러 기술들로 추출된 인간의 모습 정보는 매우 실제적인 렌더링과 얼굴의 움직임을 제공한다. Williams(1990)는 실제 얼굴의 표면에 점들을 배치하는 기술에 기반한 텍스춰 지도를 사용하였다. Mase와 Pentland(1990)는 광류법(Optical Flow)과 입의 동작을 읽기 위한 PCA(Principal Component Analysis)를 도입하였다. Terzopoulos와 Waters(1991)는 비디오 시퀀스로부터 얼굴 근육의 움직임을 예측할 수 있다고 발표했다. Kurihara와 Arai(1991)는 개별적인 얼굴의 사진을 사용하여 얼굴의 모델링과 애니메이션을 위한 새로운 변환방법을 소개하였다. Waters와 Terzopoulos(1991)는 레이저 스캐너로부터 얻어진 데이터를 사용하여 얼굴을 모델링하고 애니메이션하였다 [12]. Saji(1992)는 움직이는 간의 얼굴로부터 3차원 형태(Shape)를 추출하는 "Lighting Switch Photometry"라고 불리는 새로운 방법을 소개하였다. Kato(1992)는 얼굴 감정의 구성과 설정을 위한 Isodensity Map을 사용하였다. 이 기술은 실 시간적으로 정보를 추출하지는 않는다. 그렇지만, 대화형 입력장치로 조정되는 실시간 얼굴 애니메이션은 DeGraf (1989)에 의해 보고되었다.

3.2 컴퓨터 그래픽스에서 얼굴 애니메이션

얼굴은 사람을 인식하는데 중요하고 매우 복잡한 구조를 가지기 때문에 얼굴의 애니메이션은 컴퓨터 그래픽스 연구자들의 관심사가 되어 왔다 [9~16]. 연구자들은 얼굴의 좀 더 실제적인 애니메이션을 위하여 많은 노력을 기울여 왔다. 얼굴의 애니메이션은 일반적으로 두려움, 성냄, 놀람, 거슬림, 기쁨과 슬픔 같은 감정을 표현하는 것뿐만 아니라 화자의 입 모양 모델링으로서 인식되었다. 실제적인 얼굴의 움직임을 모델링하는 것은 정의하는 것도 어렵고 다시 구성하는 것은 더욱 더 어렵다. 왜냐하면 구성하는 것은 실제적이어야 하며, 다양한 데이터량을 가지고 능동적으로 충분히 동작할 수 있어야 하기 때문이며, 이 어려움은 주로 얼굴의 해부학적인 복잡도에 기인한다.

3.2.1 근육 구조에 기반한 얼굴의 모델링

얼굴 모델은 여러 개의 삼각형 메쉬 들로 이루

어지며, 얼굴의 중심을 관통하는 수직선에 대해서 대칭이다. 윗 입술의 꼭지점들은 아랫 입술의 꼭지점 들과 구별되도록 배치된다. 또한 이러한 꼭지점들의 구조는 계층적으로 구성할 수도 있다. 즉, 입, 눈, 코, 귀 등을 하나의 계층으로 생각하고, 그것들을 연결하는 새로운 상위 계층구조를 구성할 수도 있다. 어떤 특별한 목적을 위해서는 하나의 계층구조만으로 충분할 수 있다. 이를 테면 화자의 입 모양 만을 애니메이션하는 경우는 얼굴의 모든 근육 중에서 입의 움직임과 그 근처의 움직임에 관여하는 근육만을 고려하면 된다. 여기서 입 근처 피부 근육과 입의 근육을 여러 계층으로 만들 수도 있지만 화자의 입 모양 만을 고려하는 경우는 같은 계층으로 두고 처리하는 것이 가능하다.

애니메이션 시에는 여러 개의 삼각형 메쉬들로 이루어진 모델에서 각각의 면을 컴퓨터 그래픽스의 기술을 이용하여 렌더링을 함으로써 표현한다. 얼굴은 크게 3부분으로 나누어지게 되는데, 위 부분, 중간 부분, 아래 부분이다. 그런데 주로 움직임을 가지는 부분은 얼굴의 아래 부분에 해당하는 입을 포함하는 영역으로서 많은 움직임을 가지게 된다. 따라서 위 부분은 움직임이 없이 구성되고, 중간 부분은 아래 부분과 근육으로 연결되어 있어 아래 부분의 움직임에 따라서 영향을 받으므로 그에 따른 움직임을 형성하여야 한다. 아래 부분은 움직임을 직접적으로 받는 부분이므로 변형되고, 턱이 움직인다. 이러한 면들은 얼굴의 움직임에 따라서 주변의 움직임과 연속성을 가지고 이루어져야 자연스럽다. 따라서 얼굴을 구성하는 꼭지점들을 그 위치에 따라서 얼굴 부분의 적절한 그룹으로 분류를 해야 한다. 면 자체를 분류할 수도 있지만 어차피 애니메이션을 할 때 조절하기 쉬운 것은 꼭지점의 위치이므로 꼭지점들을 사용하여 분류한다.

얼굴에 있는 근육은 입 둘레근(Orbicularis oris)을 표현하는 4개의 직근과 왼쪽과 오른쪽 요소를 가지는 근육들의 쌍으로 구성할 수 있다. 근육의 구조는 그림 8과 같이 놓인다.

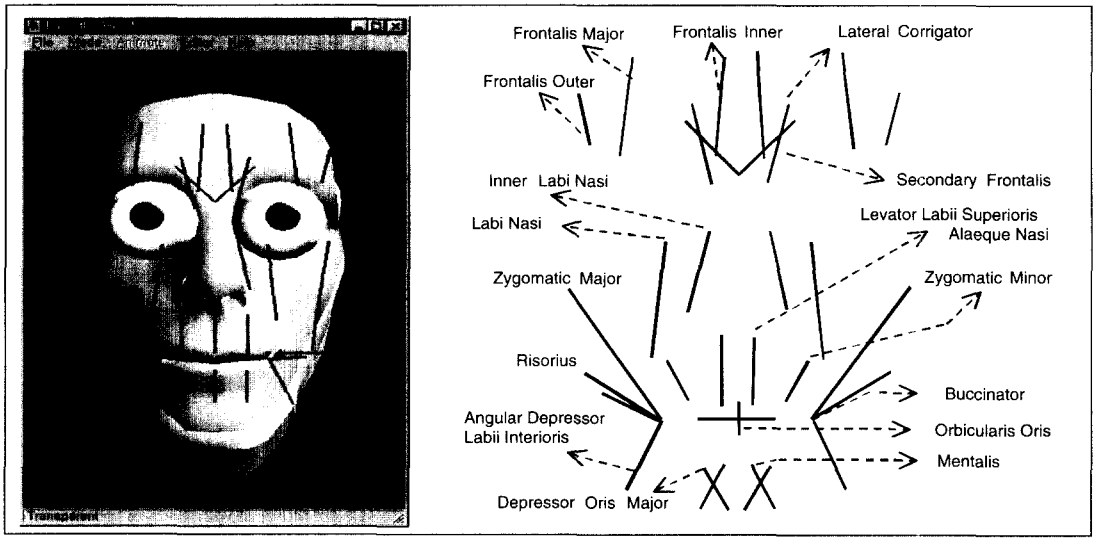


그림 8. 얼굴의 근육[11]

3.2.2 얼굴 근육의 모델링

피부는 탄력 있는 물체이다. 근육들은 피부를 변형시키는 힘으로서 모델링된다. 피부는 힘에 의해서 변형되는 다른 탄력있는 물체와 유사하다. 그렇지만 일단 변형을 이룬 후에는 더 이상의 변형은 이루어지지 않기 때문에 피부는 완벽한 탄성체는 아니다. 피부는 스프링들이 연결된 메쉬들로 모델링되며, 근육들의 장력에 의해 변형된다. 두 종류의 중요한 근육이 있다. 웃을 때 당겨지는 근육과 같은 직근(Linear)과 입 주변의 근육같이 타원형으로 오므라드는 괄약근(Sphincter)이 있다.

말을 하는데 영향을 미치는 주요한 근육들은 다음과 같다.

- *입둘레근(Orbicularis Oris)* : 이 근육은 말을 하는 동안 입의 모양을 구성하는 가장 중요한 역할을 한다. 특별히 우리가 "오" 혹은 "아"라고 할 때, 입 둘레근은 입의 모양을 결정한다. 이것은 입 주변의 근육들을 모으는 역할을 하므로 매우 중요하다.
- *턱끝근(Mentalis), 볼근(Buccinator), 입 꼬리 내림근(Depressor Anguli Oris Major), 아래 입술 내림근(Depressor Labii inferioris)* : 이 근육들은 얼굴의 아랫 부분에 위치하며 아래 입과 얼굴 아래 부분을 조정한다. 이 근육들은 입둘레근과 턱의 회전과 함께 말을 하는데 매우 중요하다.
- *작은 광대근(Zygomatic Minor), 위 입술 콧방울 올림근(Levator Labii Superioris alaeque Nasi), 위 입술 올림근(Levator Labii Superioris)* : 이 근육 들은 얼굴의 윗 부분에 위치하며 말을 하는 동안에 거의 사용되지 않는다.
- *입 꼬리 당김근(Risorius), 큰 광대근(Zygomatic Major)* : 이 근육들은 뺨 주변에 위치하며 말을 하는 것보다 감정을 표현하는데 중요한 역할을 수행한다.

3.3 MPEG-4에서 합성 얼굴의 애니메이션

MPEG-4 표준의 목적은 단지 하나의 목적에만 사용되는 이전 표준에 대하여 새로운 기술적인 경향을 제시하는 새로운 차원의 표준이다. MPEG-4는 여러 가지 종류의 데이터들을 규정하고 결합할 수 있는 유연성을 허용하여 다양한 응용목적에 사용될 수 있다. MPEG-4는 실제와 합성 음성뿐 만 아니라 영상, 3차원 그래픽까지 포함하는 진정한 멀티미디어 통신을 위한 최초의 국제 표준이다. 이 표준에 포함된 기술을 이용하여 합성 얼굴과 몸을 가지는 합성 인물을 정의하고 애니메이션할 수 있다. 머리의 예를 들자면, 입의 한 구석을 왼쪽으로 움직이는 하위 레벨의 움직임에서부터 얼굴의 감정 표현 같은 상위 레벨의 움직임까지를 정의하는 70개 이상의 모델에 독립적인 애니메이션 변수를 표준화하였다.

MPEG-4는 합성 멀티미디어 컨텐트를 고려중인데 합성음과 TTS(Text-To-Speech)를 제공한다. 합성 영상 컨텐트를 위해서는 MPEG-4는 사각형, 면, 인덱스가 붙은 얼굴 집합과 임의의 모양의 2차원 물체들과 같은 기본 구조물들로 이루어진 2차원과 3차원의 물체를 구성할 수 있도록 해준다. 3차원 물체의 정의는 VRML 노드의 하부 집합에 기반하며 2차원과 3차원 물체들의 결합을 가능하게 하도록 확장된다. 물체들은 BIFS를 사용하여 2차원과 3차원의 화면을 구성한다. BIFS(Binary Format for Scenes)는 물체의 애니메이션과 그들의 특징을 표현하는 것을 허용한다.

3.3.1 MPEG-4에서 얼굴

FBA(Face and Body Animation Ad Hoc

Group)는 인간의 얼굴과 몸체의 부호화를 다루는데, 즉, 그것들의 형태와 움직임의 표현 방법이다. 이것은 통신, 연예에서 의학분야에 이르기까지 많은 응용을 할 수 있으므로 매우 중요하다. 그러므로 표준화에 대한 강렬한 관심이 표명되고 있다. FBA 그룹은 인간의 얼굴과 몸통의 정의와 애니메이션을 위한 파라메터를 자세히 정의하였다 [17]. 이 표준안들은 가상 인간 연구의 분야를 선도하는 여러 연구소의 제안서에 바탕을 두고 있다 [9~17].

변수들의 정의는 몸통/얼굴 형태, 크기와 텍스춰의 자세한 정의를 허용한다. 애니메이션 변수들은 얼굴의 감정과 몸체의 움직임을 정의하는 것을 허용한다. 변수들은 자연적으로 가능한 모든 감정과 움직임 뿐만 아니라 만화 주인공 같은 어떤 특별한 움직임까지도 포함하고 있다. 애니메이션 변수들은 어느 얼굴/몸체 모델에 대해서도 정밀하게 구현되도록 세밀하게 정의된다.

3.3.2 FAP(Facial Animation Parameter Set)

MPEG-4는 FAP(Face Animation Parameter)의 집합을 규정한다. 각각은 그것들의 중성상태에서 얼굴 모델을 형성하는 개별적인 얼굴 움직임과 연관된다. 개별적인 FAP를 위한 FAP값은 해당되는 움직임의 크기를 나타낸다. 예를 들자면, 작은 웃음(Smile)에서 큰 웃음(Laugh)까지를 나타낼 수 있다. 개별적인 얼굴의 움직임 시퀀스는 해당되는 시간에 규정된 FAP 값에 따라서 중성 상태에서 얼굴 모델을 변형시킨다. FAP들은 최소한의 얼굴의 움직임의 연구에 바탕을 두고 있으며 근육의 움직임과 연관되어 있다. 이것들은 기본적인 얼굴의 동작의 완전한 집합을 표현하며, 따라서 가장 자연스러운 얼굴 감정의 표현을 허용한다.

변이 움직임을 가진 모든 변수들은 FAPU (Facial Animation Parameter Units)의 항으로 표현된다. 이 단위들은 감정과 말의 발음에 대한 그럴듯한 결과를 나타내기 위하여 동일한 방법으로 얼굴 모델에 대한 FAP들의 확장해석을 허용하도록 정의된다. 이것들은 예를 들자면 눈사이의 거리와 같은 중요한 얼굴 특징점(Key Facial Feature)간의 거리의 비율이다. 비율의 단위는 충분한 정밀도를 허용하도록 선택된다.

변수의 집합은 두가지의 상위 레벨의 변수들을 제공한다. Viseme 변수는 다른 변수들의 사용없이도 얼굴의 Viseme들을 렌더링하는 것을 허용한다. 비슷하게 Expression 변수는 얼굴의 상위 레벨 감정 표현의 정의를 허용한다.

표 1과 같이 68개의 FAP 변수들은 얼굴의 부분에 대해서 서로 관련이 있는 10개의 그룹으로 구분되어진다.

표 1. FAP 그룹

1: visemes and expressions	2
2: jaw, chin, inner lowerlip, cornerlips, cornerlips, midlip	16
3: eyeballs, pupils, eyelids	12
4: eyebrow	8
5: cheeks	4
6: tongue	5
7: head rotation	3
8: outer lip positions	10
9: nose	4
10: ears	4

3.3.3 FDP(Facial Definition Parameter set)

얼굴의 애니메이션을 제공하는 MPEG-4 복호화기는 FAP들을 해석하고 표현할 수 있는 일반적인 얼굴 모델을 가지고 있어야 한다. 이것은

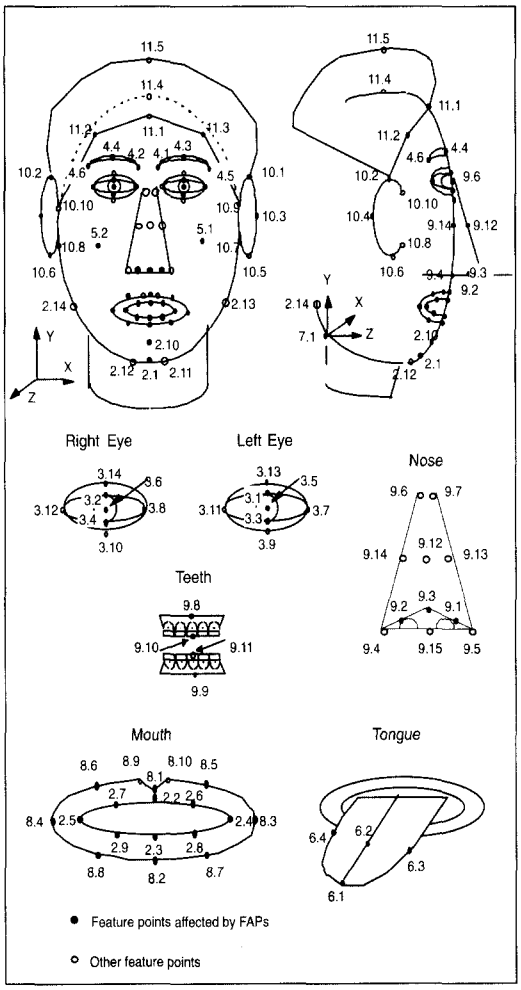


그림 9. 얼굴 모델을 정의하기 위해 사용되는 특징점들[17]

얼굴 감정 표현과 발음하는 모양을 재 생성할 수 있어야 함을 의미한다. 또한 일부분을 수정함에 의해서 특정한 인물/주인공같이 보이도록 하기 위해 FDP가 필요하다. MPEG-4는 얼굴 애니메이션 파라메터들을 정의하기 위하여 참조되도록 그림 9와 같이 얼굴 부분에 위치하는 84개의 특징점(Feature)을 규정한다. 머리 선에 따라서 위치하는 어떤 특징점들은 FAP들에 의해서 영향을 받지 않는다. 다만 특징점들을 사용하여 특정한 얼굴 모델의 모양을 정의할 때 필요하다. 특징점들은 뺨, 눈, 입과 같은 그룹들로서 정리되어 있다.

FDP들은 특별한 얼굴의 일반적인 얼굴 모델에 개인적인 특성을 주기 위하여 사용된다. FDP 집합은 3차원 특징점, 특징점에 대한 텍스춰 좌표계, Face Scene 그래프, FAT(Face Animation Table)과 같은 것을 포함할 수 있다. 텍스춰 좌표계는 각 특징점들에 대해 제공되며, Face Scene 그래프는 물질적인 특질 뿐만 아니라, 다중 면(Surface)과 텍스춰들을 포함하는 3차원 다각형 모델(Polygon Model)이다. FAT(Face Animation Table)는 어떻게 얼굴이 부분적 선형 함수로서 각 FAP에 대한 Face Scene 그래프에서 꼭지점들의 움직임을 규정함에 의해서 애니메이션될 수 있는지를 규정한다.

3.4 실감나는 가상 인물의 렌더링

실제처럼 보이는 가상 인간(혹은 합성 배우라고도 불리는 것)을 개발하는 중요한 이유는 실제 세계를 표현하는 어떠한 가상 화면에서도 이용할 수 있기 때문이다. 가상 화면이 아무리 아름다워도 인간 없이는 완전하지 않다. 가상 인간들을 가진 화면들은 매우 복잡한 문제를 가지고 있다. 실제처럼 보이도록 가상 인물을 표현하기 위해서는 몸통, 얼굴과 의복의 변형까지 고려해야 한다. 또한 실제적인 가상 인물의 구축은 관심있는 인물의 특별한 특징점을 포함시켜야 한다. 예를 들자면 마릴린 몬로, 험프리 보가드, 엘비스 프레슬리같은 인물의 개인적인 특성을 포함시켜야 한다. 이러한 인물들은 약간의 차이가 있더라

도 관찰자들은 그것을 쉽게 인식할 수 있다. 가까운 미래에는 우리는 가상 인물과 실제 인물의 구분을 할 수 없을 지도 모른다. 그리고 좀 더 자연스럽게 가상 환경을 실시간 적으로 보여줄 수 있을 것이다.

3.4.1 피부 텍스춰

가상 인물의 질감을 향상시키기 위하여 실제 얼굴의 사진을 텍스춰 매핑시키는 기술을 사용한다. 주어진 사진의 3차원 얼굴 구조를 일치시키기 위한 도구로서 텍스춰 매핑이 개발되었다. 완전한 일치를 위해서 3차원 모델의 특징점 중에서 사진의 부분과 일치하는 몇 개의 점이 선택된다. Delaunay 삼각형은 이러한 점들을 연결하기 위하여 사용된다. 이 점들은 능동적으로 그림위에 놓인다. 삼각형 영역(Triangular Domain)에서 보간(Interpolation) 구조는 원하는 텍스춰 좌표를 얻기 위하여 사용된다. 결과적으로 그림은 3차원 모델 위에 변형되어 배치된다. 완전한 머리의 배치를 위하여 다양한 각도에서 바라본 정보가 필요하다. 이 그림들은 원통에 투영된 후, 해당되는 배치가 3차원 원통에 투영된 3차원 모델의 점들과 3차원 원통에 투영된 그림간에서 이루어진다. 텍스춰 매핑을 사용하여 렌더링 성능은 눈에 띄게 향상된다. 덧붙여서, 주어진 3차원 모델에 특정한 인물의 사진을 올려놓는 것을 가능하게 했다.

색과 밝기에 따라서 변화하는 여러가지 타입의 피부들의 가상 텍스춰들을 생성할 수 있다는 것은 매우 흥미롭고 유용하다. 얼굴의 서로 다른 여러 영역에 대한 피부의 표본들은 가상 피부 텍스춰를 생성하는데 사용한다. 영상을 다루는 기술은 그룹에서 표본들에 대한 특징점들

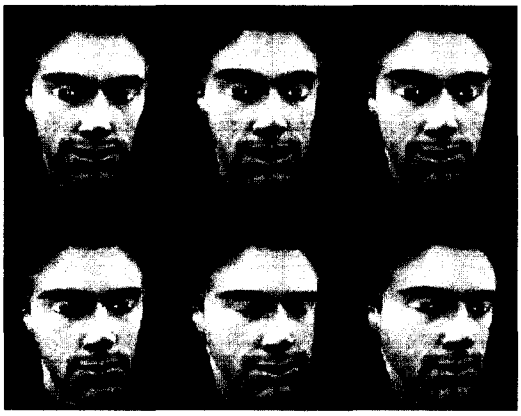


그림 10. 피부 텍스춰 렌더링의 예[15]

을 얻기 위해 사용되며, 각 그룹들은 그림 10과 같이 피부 텍스춰에 맞는 피부의 부분에 놓여진다.

3.4.2 머리카락 모델링(Hair Modeling)

인간의 시뮬레이션에서, 머리카락은 가장 어렵고 도전할 만한 분야이다. 이것은 지금까지 만들어진 기술 중에서 가상 인간을 만드는데 만족할 만한 성능을 얻지 못한 분야중의 하나이다. 머리카락 프로세싱의 어려움은 머리카락의 표현은 매우 많은 수의 세밀한 기하학적 구조인 머리카락의 빛과 어둠에 따른 음영처리와 렌더링된 영상과 비교해 볼 때 매우 가는 조그만 크기인 물체를 처리하기 때문에 발생한다. 이를 위하여는 기본적으로 다음과 같은 단계를 거친다. 먼저 머리카락들에 대한 데이터베이스를 구축(좌표값, 색상, Normal 벡터 성분등)하고 모든 광원에 대한 그림자를 계산하고 생성한다. 여기에 그림자들을 이용하여 머리카락이 없는 물체를 먼저 구성한 후, 머리카락을 추가하는 과정을 거친다.

그림 11. 헤어 스타일의 렌더링[13]

더 나아가서, 헤어 스타일 렌더링에서는 다음과 같은 과정을 거친다. 먼저 장면의 음영을 각 광원으로부터 계산하고, 물론 머리카락의 음영부분도 계산한다. 머리카락의 음영부분은 물체의 표면과 개별적인 머리카락 각각에 대하여 계산된다. 마지막으로, 헤어 스타일이 장면에 따라서 구성된다. 또한, 물이나 바람과 같은 것에 의해 변형되는 것도 다루어야 한다. 그림 11은 이러한 기술들이 종합된 한 예이다.

3.4.3 의복 모델링(Clothes Modeling)

가상 인물을 도입하기 위하여 가장 최근의 컴퓨터를 사용해 생성된 영화들에서는 복장은 몸체의 한 부분으로서 취급된다. 좀 더 실제적인 의복의 모델링을 위하여, 두 가지의 문제가 먼저 해결되어야 한다. 충돌부분(Collision Detection)이 없는 의복의 움직임과 몸체위에 입혀진 의복의 충돌부분 처리이다. 유연하고 변형되는 물체는 하나의 물체로 간주할 수 없고, 그것의 움직임이 작은 수의 몇 개의 점으로부터 계산할 수

그림 12. 의복 모델링의 예[13]

없기 때문에 견고한 물체(Rigid Object)와 다르다. 유연한 물체는 작은 부분으로 나누어져야 하고 각 점들은 지역적인 제약과 전역적인 제약에 따라 배치된다. 이러한 제약들은 이러한 제약들간의 충돌을 방지하도록 하는 힘을 생성한다. 동적 시스템을 푸는 것은 모든 힘들간의 동질성(Equilibrium)을 찾는 것을 요구한다. 최근의 연구는 유연하고 변형되는 물체에 대한 동적 모델을 다룬다.

4. 응용 분야

얼굴 애니메이션은 많은 분야에 응용이 될 수 있는데, 가상회의 시스템에서 자신을 대표하는 물체인 아바타(Avatar)의 얼굴 움직임에 적용될 수 있다. 초저속 채널을 이용해서 충분한 정보를 전달할 수 있다는 특성 때문에 이동 통신체에서의 통신에 사용될 수 있다. 이러한 분야를 간략히 정리하면 그림 13과 같다.

얼굴 애니메이션의 응용 분야를 정리하자면 가장 대표적인 응용 분야는 영상통신이다. TTS

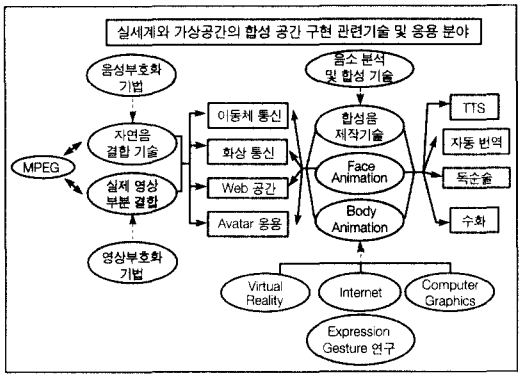


그림 13. 실세계와 가상 공간의 합성공간 구현 관련 기술 및 응용분야

는 문자로 쓰여 있는 문장을 합성음으로 만들어 내는 방법으로 이 기술에 대한 부분은 이미 MPEG-4 SNHC부분에 포함되어 있다. 입술 모양을 보고 상대방이 말하는 것을 이해하는 기술인 독순술도 한 응용 예이다. 맹인들의 경우는 점자를 가지고 상대방이 말한 바를 이해할 수 있다. 그런데 귀머거리인 경우는 말을 들을 수 없기 때문에 필담만을 나눌 수 있다. 그런데, 독순술이란 기술은 상대방이 말할 때의 입술 모양을 보고서 상대방의 말하는 바를 알 수 있다. 물론 수화를 사용하기도 하지만 이것이 좀 더 자연스럽기도 하다.

성형수술 분야에서는 다음과 같이 응용될 수 있다. 얼굴 부분의 변수를 미리 읽은 후 성형 수술 후 새롭게 변수가 변화하면 어떤 얼굴형을 가지게 될 것인지를 예상하는데 사용될 수 있다. 성형수술을 하려는 환자는 수술 후 자신의 모습이 어떻게 될 건지에 많은 관심을 가지고 있다. 이것을 아는 방법으로는 컴퓨터 시뮬레이션을 통하여 알 수도 있고, 직접 사진에 해당 부위를 처리한 후 그 변화되는 모습을 보는 방법도 있다. 그렇지만 그것은 시간을 많이 요구한다.

패션 분야에서는 모델이 화장을 하거나 의상을 입으면 어떤 이미지를 주는지를 미리 예측하는데 응용될 수 있다.

그림 13에서 보듯이 이 분야의 관련 기술 및 응용 분야는 엄청나게 넓다는 것을 알 수 있다. 정상인에게는 합성공간에서의 실제감을 제공하며, 신체 부자유자에게는 보다 편리한 시스템을 제공함으로써 신체적인 제한이 작업에 지장을 주지 않도록 하는 효과를 얻을 수 있을 것이다. 또한 개인이 새로운 프로그램을 만드는 창조적인 작업에 참여하는 기쁨을 제공해줄 것이다. 이러한 여러 분야의 결합은 해당 분야에 많은 고용창출도 발생시킬 것으로 기대된다.

감사의 글

본 연구는 광주과학기술원(K-JIST) 초고속광네트워크 연구센터(UFON)를 통한 한국과학재단 우수연구센터(ERC)와 교육부 두뇌한국21(BK21) 정보기술사업단의 지원에 의한 것입니다.

참 고 문 헌

[1] T. Kanade and M. Okutomi, "A Stereo Matching Algorithm with an Adaptive Window: Theory and Experiment," IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 16, no. 9, Sept. 1994.
[2] D. Scharstein and R. Szeliski, "Stereo Matching with Non-Linear Diffusion," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 343-350, June 1996.
[3] R. Szeliski, "Stereo Algorithms and Representations for Image-Based Rendering," 10th British Machine Vision Conference, vol. 2, pp. 314-328, Sept. 1999.
[4] 이범구, 김희정, 박성춘, 남승진, "가상현실과 방송응용," 방송공학회지, 제4권, 제3호, pp. 230-238, 1999년 9월.
[5] 박남준, 이재호, 권영무, 박상희, "중간영상 합성을 위한 다해상도 스테레오 정합기법," 방송공학회논문지, 제2권, 제2호, pp. 216-224, 1997년 12월.
[6] F. Devernay and O. Faugeras, "Computing Differential Properties of 3-D Shapes from Stereoscopic Images without 3-D Models," Research Report RR-2304. INRIA, Sophia Antipolis, 1994.
[7] T. Kanade, K. Oda, A. Yoshida, M. Tanaka, and H. Kano, "Video-Rate Z Keying:A New Method for Merging Images," Tech. Report CMU-RI-TR-95-38, Robotics Institute, Carnegie Mellon University, Dec., 1995.
[8] K. Kutulakos and J. Vallino, "Calibration-Free Augment Reality," IEEE Trans. Visualization and Computer Graphics, vol. 4, no. 1, pp. 1-20, 1998.
[9] F. I. Parke and K. Waters, Computer Facial Animation, A K Peters, Wellesley, 1996.
[10] P. Ekman, Darwin and Facial Expressions, Academic Press, New York, 1973.
[11] K. Waters, "A Muscle Model for Animating 3D Facial Expressions," SIGGRAPH' 87 Computer Graphics, vol. 21, pp. 17-24, 1987.
[12] D. Terzopoulos and K. Waters, "Analysis and Synthesis of Facial Image Sequences using Physical and Anatomical Models," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.15, pp. 569-579, 1993.
[13] N. Thalmann and D. Thalmann, Computer Animation: Theory and Practice, Springer-Verlag, 1990.
[14] S. Tu and D. Terzopoulos, "Perceptual Modeling for The Behavioural Animation of Fishes," Proc. Second Pacific Conf. On Computer Graphics, 1994.
[15] A. Watt and M. Watt, Advanced Animation and Rendering Techniques, Addison-Wesley, 1992.
[16] A. Watt, The Computer Image, Addison-Wesley, 1998.
[17] P. van Beek, E. Petajan, and J. Ostermann, "MPEG-4 Synthetic Video," in Multimedia Systems, Standards, and Networks, Marcel Dekker, pp. 299-330, 2000.

필 자 소 개

전 정 희

- 1993년 2월 : 조선대학교 컴퓨터공학과 (공학사)
- 1997년 8월 : 조선대학교 컴퓨터공학과 (공학석사)
- 2001년 2월 : 조선대학교 컴퓨터공학과 (공학박사)
- 2001년 2월~현재 : 광주과학기술원 Post doctor
- 관심분야 : 3차원 영상합성, 스테레오 비젼, 컴퓨터그래픽스

필 자 소 개

임 동 근

- 1994년 2월 : 전북대학교 전자공학과 (공학사)
- 1993년 11월~1995년 2월 : 현대전자(주) 반도체 제2연구소 ASIC분야(연구원)
- 1997년 2월 : 광주과학기술원 정보통신공학과 (공학석사)
- 1997년 3월~현재 : 광주과학기술원 정보통신공학과(박사과정)
- 주관심분야 : 영상신호처리, 동영상 부호화 및 컴퓨터그래픽스, 고속 VLSI 회로설계

호 요 성

- 1981년 2월 : 서울대학교 전자공학과 (학사)
- 1983년 2월 : 서울대학교 전자공학과 (석사)
- 1983년 3월~1995년 9월 : 한국전자통신연구소 선임연구원
- 1989년 12월 : University of California, Santa Barbara Department of Electrical and Computer Engineering (박사)
- 1990년 1월~1993년 5월 : 미국 Philips 연구소 Senior Research Member
- 1995년 9월~현재 : 광주과학기술원 정보통신공학과 부교수
- 주관심분야 : 디지털 신호처리, 영상신호처리 및 압축, 초저속영상통신, 디지털 TV와 고선명 TV방식, 멀티미디어 통신방식