

# Facial Animation을 위한 다중 마커의 추적

°이 문희\*, 김 철기\*\*, 김 경석\*\*  
\*부산대학교 일반대학원 멀티미디어학과  
\*\*부산대학교 전자계산학과

## The Multi-marker Tracking for Facial Animation

Mun-Hee Lee\*, Cheol-Ki Kim\*\*, Kyong-Sok Kim\*\*  
\*Dept of Multimedia, Pusan National University  
\*\*Dept of Computer Science, Pusan National University

### 요약

얼굴 표정을 애니메이션하는 것은 얼굴 구조의 복잡성과 얼굴 표면의 섬세한 움직임으로 인해 컴퓨터 애니메이션 분야에서 가장 어려운 분야로 인식되고 있다. 최근 3D 애니메이션, 영화 특수효과 그리고 게임 제작 시 모션 캡처 시스템(Motion Capture System)을 통하여 실제 인간의 동작 및 얼굴 표정을 수치적으로 측정해내어 이를 실제 애니메이션에 직접 사용함으로써 막대한 작업시간 및 인력 그리고 자원을 획기적으로 줄이고 있다. 그러나 기존의 모션 캡처 시스템은 고속 카메라를 이용함으로써 가격이 고가이고 움직임 추적에서도 여러 가지 문제점을 가지고 있다. 본 논문에서는 일반 저가의 카메라와 신경회로망 및 영상처리기술을 이용하여 얼굴 애니메이션용 모션 캡처 시스템에 적용할 수 있는 경제적이고 효율적인 얼굴 움직임 추적기법을 제안한다.

### 1. 서론

최근, 모션 캡처 시스템은 영화, 3D 애니메이션, 게임 그리고 광고를 비롯한 방송 콘텐츠 등의 고부가가치 엔터테인먼트 산업에서 가장 활발하게 활용이 되고 있다. 또한 의료분야 및 스포츠·레저, 각종 기초과학 및 공학분야에서도 첨단분석장비 및 실험장비로 폭넓게 활용되고 있으며 점점 그 적용범위를 확대해 가고 있는 실정이다[1][10].

모션 캡처 시스템은 어떤 특정분야에 한정된 시스템이라기보다는 멀티미디어 콘텐츠 제작도구 및 첨단분석도구로써 여러 분야에서 활용될 수 있을 것으로 예상되며 관련시장은 기하급수적으로 증가하고 있다[1][3]. 따라서 모션 캡처 시스템에 대한 많은 관심과 시급한 연구·개발이 요구되어지며 특히, 광학방식 모션 캡처 시스템의 연구 분야 중 하나가 얼굴을 애니메이션 하기 위한 얼굴 움직임 추적기법으로, 이는 모션 캡처 시스템에서 가장 핵심적인 기술이라 할 수 있다. 사람의 신체에서 얼굴이 차지하는 비율은 적지만, 사람을 애니메이션 하는데 있어서 가장 기본적인 요소이다. 그러나 얼굴이 가지고 있는 구조의 복잡성으로 인하여 만족할만한 연구 결과가 나오지 못하고 있다[7].

본 연구에서는 고가 고속의 카메라를 이용한 기존의 모션 캡처 시스템에 비해 일반 저가의 카메라를 이용하여 얼굴 움직임을 빠르고 효과적으로 추적할 수 있는 얼굴 애니메이션용 광학방식 모션 캡처 시스템 기술을 제안하고자 한다.

### 2. 관련연구

기존의 얼굴 애니메이션 기법은 다음과 같다.

#### 2.1 키프레임(Keyframe) 애니메이션 기법

얼굴의 움직임을 만들기 위해 주요한 키프레임들을 정하고 인 비트윈(In-between)기법을 적용하여 키프레임들 사이를 부드럽게 연결시키는 애니메이션 기법이다. 이 방식은 2차원에서 사용되었던 기법을 3차원에 그대로 적용한 것이다. 그러나 2차원에서와는 달리 3차원에서는 2차원보다 얼굴 움직임을 위한 많은 키프레임들을 필요로 하기 때문에 많은 양의 수 작업이 필요하다. 따라서 이에 따른 작업시간, 인력 그리고 엄청난 자본이 요구된다는 특징을 가지고 있다[1][7].

#### 2.2 가상근육기반(Pseudomuscle Based)애니메이션 기법

얼굴의 상하구조는 피부와 근육, 뼈로 나누어질 수 있으며, 근육의 움직임에 의하여 얼굴의 다양한 표정이 나타난다. 가상 근육기반 애니메이션 기법이란 얼굴의 정확한 해부학적 구조를 시뮬레이션하는 것이 아니라 몇 개의 제어점에 의하여 기본적인 얼굴 근육의 움직임을 나타내는 것이다. 그러나 이 방법은 근육을 시뮬레이션하는 것보다 간단하지만 근육과 표정의 움직임이 일치하지 않기 때문에 3차원의 구조를 갖는다고 보기 어렵다[7][9].

2.3 근육기반(muscle based)의 애니메이션 기법

근육기반(muscle based)의 모델링 기법은 얼굴의 해부학적인 구조에 기반을 둔 기법이다. 이 방법은 얼굴 근육과 피부의 특성을 시뮬레이션하여 얼굴 표정을 나타낸다. 현재까지 얼굴의 복잡한 구조를 정확하게 반영한 얼굴 애니메이션 기법은 없으나, 뼈와 근육 구조를 단순화하여 모델링한 기법들이 제시되어 있다[7].

2.4 광학방식 모션 캡처(Optical Motion Capture)

실제 배우(actor)의 얼굴에 관찰하고자 하는 관찰영역에 적외선 반사체(infrared Marker) 및 자기센서(magnetic sensor)를 부착하고 반사체 및 자기센서의 정보를 컴퓨터로 전송하여 얼굴의 관찰 영역 움직임을 포착하고, 생성된 움직임 정보를 3차원 애니메이션 소프트웨어 및 분석 소프트웨어와 링크(link)함으로써 실제 배우와 같은 섬세하고 유연한 동작의 고품위 3D 애니메이션 제작을 가능하게 하는 애니메이션 기법이다[1][3].

모션 캡처 시스템은 기계(Mechanic), 자기(Magnetic) 그리고 광학(Optical)방식으로 구분할 수 있다. 기계방식은 배우의 각 관절에 슬라이더 엔코더(Slider Encoder)를 부착하여 컴퓨터로 각도를 전달하는 방식이다. 기계장치를 배우의 신체에 부착해야 되기 때문에 행동에 커다란 제약과 가지고 있다.

자기방식은 배우의 각 관절에 여러 개의 자기센서를 부착하고 유선 또는 무선으로 컴퓨터에 연결하여 움직임을 캡처하는 방식이다. 비교적 가격이 저렴한 반면, 자기센서의 개수가 제한적이다. 광학방식은 적외선 반사체(마커)를 배우의 각 관절에 부착하고, 6~32대 가량의 적외선 카메라를 이용하여 적외선에 반사되는 적외선 마커를 여러 장의 2D 이미지로 획득하여 3D 공간상의 좌표로 움직임을 파악하는 방식이며 다른 방법들 보다 섬세하고 유연하며 빠른 동작에서도 움직임 정보를 잘 감지할 수 있다. 기존의 광학방식 모션 캡처 시스템은 고속 카메라와 적외선 조명을 사용함으로써 매우 고가로 판매되고 있다.

얼굴용(Facial) 모션 캡처 시스템의 경우 대개 광학방식을 사용하며, 배우의 얼굴에 20~30개의 마커를 부착한 후 필요한 정밀도에 따라 마커의 수를 조절하여 사용한다. 카메라는 보통 1~2대를 사용하는 것이 일반적이다.

3. 얼굴 움직임의 추출 및 추적 기법

실제 사람의 얼굴에 다수 마커들을 부착하고 얼굴 움직임 추출 전용 헤드셋(headset)[그림 2]을 이용하여 영상을 입력하고, Optical flow에 의한 지역 움직임(local motion) 분석 방법을 적용하여 다중 마커들을 추출한후, 신경회로망을 이용하여 다중 마커들을 추적한다.

3.1 Optical flow에 의한 지역 움직임(local motion) 분석을 이용한 다중 마커 추출

본 연구에서는 비교적 배경과 마커를 분리하기가 용이하며 노이즈가 발생하는 입력영상에 대해 계산량이 적고 실시간 다중 마커 추출이 용이하며 동영상 해석에 많이 사용되는 Optical flow에 의한 지역 움직임(local motion) 분석을 적용하여 마커를 추출해낸다. 이때 각각의 마커에 대한 히스토리(history)를 생성한다. 마커의 히스토리란 추출되는 마커의 현재 위치, 중심점, 크기 등의 특징을 의미한다. 이러한 마커의 히스토리는 레이블링과 마커 추적에서 유용하게 이용이 된다. 그리고 입력 영상을 획득하기 위한 카메라의 위치에 따라서 마커추출의 결과는 달라진다[8]. 따라서 본 연구에서는 얼굴을 정면으로 주시하는 방향으로 카메라가 부착된 헤드셋(head set)[그림 2]을 자체 제작하여 착용함으로써

써 마커들 사이의 간격을 좀더 분명하고 정확하게 구분을 지을 수 있었기 때문에 마커를 더 잘 추출할 수 있었고, 전반적인 영상 처리를 효율적으로 할 수 있었다. 또한 기존의 여러 추출 기법과는 달리 마커의 이동 속도가 작기 때문에 계산이 간단하고 효과적이며 처리시간과 성능면에서도 신경회로망을 이용한 영역 분할 및 병합기법 못지 않는 뛰어난 결과를 나타내었다[2][4][12].

영상에서의 움직임(motion)에는 두 가지 종류가 있다. 카메라의 움직임에 의해 발생하는 전역 움직임(global motion)과 움직이는 마커에 의해 발생하는 지역 움직임(local motion)이다. 움직이는 마커를 추출하기 위해서는 전역 움직임의 보상이 선행되어야 한다. 전역 움직임이 보상되면 영상에서의 움직임은 마커의 움직임에 의해 발생하는 것이다. 본 논문에서는 이러한 전역 움직임은 없는 것으로 가정한다. 왜냐하면, 본 실험의 환경이 전용 스튜디오에서 실행되었기 때문이다.

움직이는 마커에 의한 지역 움직임은 Optical flow에 의해 추정 될 수 있다. Optical flow란 움직이는 마커에 의해 형성되는 광도패턴(brightness pattern)의 이동을 의미한다. 이는 마커의 이동에 의해 영상에서 velocity field로 나타낼 수 있다[12][13][14].

시간 t에서, 영상에서의 한 점 (x,y)의 intensity를 I(x,y,t)라고 표현한다. 영상간의 움직임이 적다고 가정하면, 연속영상에 대하여 intensity I(x,y,t)의 변화율은 0가 되어 (1)식과 같이 나타낼 수 있다.

$$\frac{d}{dt} I(x, y, t) = 0 \tag{1}$$

(1)식은 chain rule에 의해 (2)식과 같이 된다.

$$\frac{\partial I}{\partial x} \frac{dx}{dt} + \frac{\partial I}{\partial y} \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0 \tag{2}$$

여기서 I=I(x,y,t) 이다.

$$U = \frac{dx}{dt}, v = \frac{dy}{dt} \text{ 라고 하면, (3)식과 같이 된다.}$$

$$I_x U + I_y V + I_t = 0 \tag{3}$$

여기서  $I_x = \frac{\partial I}{\partial x}, I_y = \frac{\partial I}{\partial y}, I_t = \frac{\partial I}{\partial t}$  이고, u와 v는 optical flow vector를 이루는 구성성분들로서 velocity vector들이다.

(3)식에서 u, v를 계산하기 위해 많은 방법들이 제안되었는데 Lucas와 Kanade가 제안한 방법이 빠른 속도를 나타낸다. Lucas와 Kanade는 작은 영역B에 소속된 픽셀에 대한 움직임 벡터는 동일하다는 가정하에 u, v를 계산하였다.

영역 B에 대한 optical flow 방정식의 오차 E는 (4)식과 같이 정의될 수 있다.

$$E = \sum_{x \in B} (I_x u + I_y v + I_t)^2 \tag{4}$$

(4)식을 x,y에 대하여 편미분하면 (5), (6)식과 같이 된다.

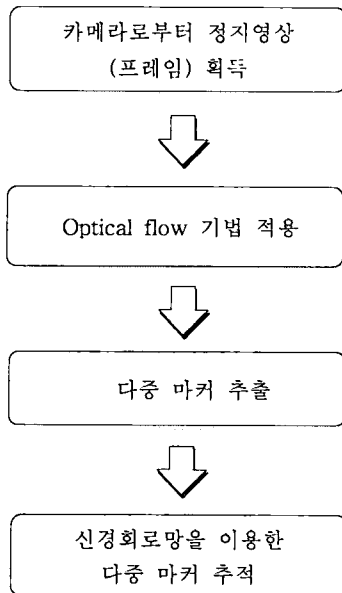
$$\sum_{x \in B} (I_x u + I_y v + I_t) I_x = 0 \tag{5}$$

$$\sum_{x^t \in B} (I_x u + I_y v + I_t) I_y = 0 \quad (6)$$

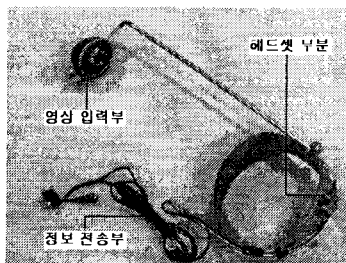
(5), (6)식을 계산하여 해를 구하면 (7)식과 같이 된다.

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \sum_{x^t \in B} I_x I_x & \sum_{x^t \in B} I_x I_y \\ \sum_{x^t \in B} I_x I_y & \sum_{x^t \in B} I_y I_y \end{bmatrix}^{-1} \times \begin{bmatrix} - \sum_{x^t \in B} I_x I_t \\ - \sum_{x^t \in B} I_y I_t \end{bmatrix} \quad (7)$$

(7)식에서와 같이 움직임이 있는 픽셀에 대해서는 속도벡터가 존재하게 된다[12][13].



[그림 1] Facial Optical Motion Capture System 흐름도



[그림 2] Facial Motion Capture System의 Head Set

### 3.2 신경회로망을 이용한 다중 마커 추적

마커의 추적은 연속적인 프레임들 속에서 수행되는 단계이다. 단일 프레임에서 추출된 마커들을 각각 저장하고 계속적으로 연결되는 프레임에서 추출된 마커들과 저장된 마커들을 여러 가지 추출된 정보를 이용하여 관련시키고 예상하여 추적하는 단계이다.

기존의 추적기법으로는 크게 중심점 추적기법과 상관 추적기법이 있다. 중심점 추적기법은 이동 표적을 배경으로부터 분리한 다음, 추출된 표적 영역의 중심점을 추적하는 방식이고, 상관 추적기법은 현재 프레임의 이동 표적의 위치에 적당한 크기의 영역을 정의하고, 정의된 영역과 다음 프레임의 검색 영역과의 상관도를 계산하여 그 상관도가 가장 높은 영역으로 표적의 이동을 추정하는 기법이다[4][11].

본 논문에서는 마커의 추적을 위해서 상관 추적이 기반하고 인공신경망을 이용한 다중 마커 추적 알고리즘을 사용하였다. 추출된 마커들로부터 이전 프레임의 마커들과의 중심점의 위치, 각 마커들의 운동벡터, 마커들간의 상대적 위치 관계등을 고려하여 상관도를 계산한 후 정합 하는 방법을 통해 추적을 수행하였다. 마커 추적 단계를 수행한후 마커의 히스토리에 현재 마커의 수정된 벡터, 크기, 그리고 위치들을 갱신한다.

[알고리즘] 신경회로망을 이용한 다중 마커 추적 알고리즘  
각 프레임에 대해서 STEP 1 ~ STEP 6을 반복 수행

STEP 1: 카메라 입력 동영상에서 추출된 마커를 입력패턴으로한다.

STEP 2: 각 마커들에 대해 전후 프레임의 상관도를 계산하여 승자뉴런을 찾는다.

STEP 3: 정합 경계값과 비교, 만약에 경계 테스트를 통과 하면 STEP 4로 실패하면 STEP 5로 간다.

STEP 4: 입력 패턴에 따른 뉴런의 가중치를 조정한다.

STEP 5: 새로운 마커 뉴런을 생성한다.

STEP 6: 다중 마커의 추적

## 4. 실험결과 및 분석

### 4.1 실험환경

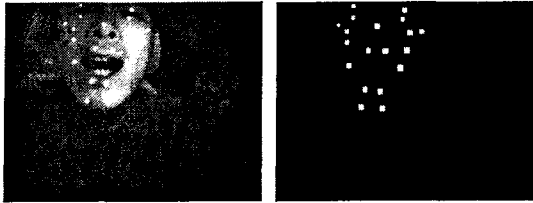
시스템 사양은 Pentium III 550MHz, RAM 128MB, Win98에서 Visual C++ 6.0을 이용하여 실험하였다. 실험에 사용한 카메라는 삼성 CCTV B/W CAMERA SBC 341이며 프레임 그래버는 MuTech사의 Image/VGA 450pro Frame grabber를 사용하여 320×240 정지영상(프레임)을 초당 15 프레임 내외로 처리하였다. 1대의 카메라를 이용하여 얼굴 정면의 표적과 좌우의 고개 움직임을 관찰하였다. 실제 얼굴에 부착한 마커는 19개를 사용하였다.

### 4.2 결과 및 분석

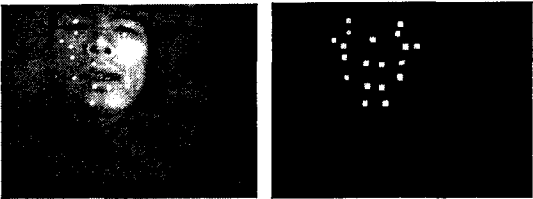
#### 4.2.1 Optical flow에 의한 지역 움직임(local motion) 분석을 이용한 다중 마커 추출 결과

본 논문에서 제시한 Optical flow에 의한 지역 움직임(local motion) 분석을 적용한 결과, [그림 3]에서와 같이 카메라 광각 내에 포착된 모든 마커에 대해서 빠른 시간 내에 다중 마커들을

추출함으로 실시간 처리에 적합한 좋은 결과를 나타내었다.



(가) 추출 결과 I



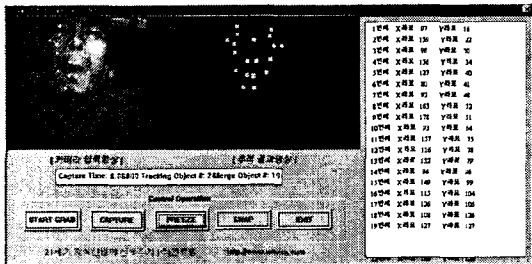
(나) 추출 결과 II

[그림 3] Optical flow에 의한 지역 움직임 분석 결과

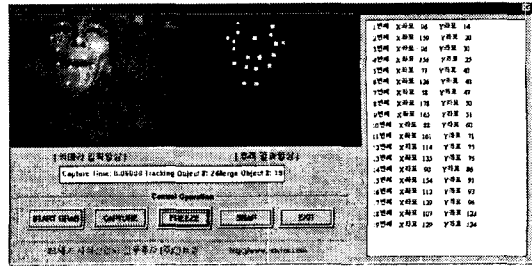
4.2.2 신경회로망을 이용한 다중 마커 추적 실험 결과

이마, 눈, 볼 그리고 턱 부분의 경우 제시한 추적기법을 적용하여 잘 추적을 할 수 있었다[그림 4]. 이는 마커의 변화량이 입술 부위와는 달리 대체로 일정하기 때문에 프레임간의 추적이 잘 된다.

그러나 입술부위의 경우, 다른 부위에 비해 움직임의 다양한 변화가 급격하게 발생하는 특징을 가지고 있다. 이러한 급격한 큰 변화량이 발생한 경우, 입술부위에 대해 마커 추적이 실패할 가능성이 높다. 따라서 본 논문에서는 추적에 실패한 부분에 대해서 경계값을 변화하여 부가적인 추적을 함으로써 그에 대한 정확도를 높일수 있었다. 얼굴의 움직임 추적은 신체의 다른 부위 보다는 공간적 시간적 변화량이 적고 부분별로 다른 특징을 가지고 있기 때문에 좀더 섬세하고 완벽한 추적기법을 요구한다. 본 논문에서는 기존의 Target tracking 기술[5][6]을 애니메이션 자동 생성기구인 모션캡처시스템에 적용해보았으며, 이를 위하여 앞서 제시한 신경회로망기반의 추적 기법을 적용해 본 결과 효과적인 결과를 얻었다. [그림 4]은 시간 t 일 때 마커들의 위치와 t+1일 때 마커들의 위치를 나타내었으며, 각 시간대의 위치벡터를 통하여 추적의 결과를 나타내고 있다



(가) 시간 t일 때 마커들의 위치



(나) 시간 t+1일 때 마커들의 위치  
[그림 4] 신경회로망을 이용한 추적기법의 결과

5.결론

본 논문에서는 얼굴 표정 애니메이션을 위해서 얼굴용 광학방식 모션 캡처 시스템 개발을 위한 효과적인 추출 및 추적기법을 제시하였다. 비교적 마커와 배경과의 대조가 잘 나타나고 노이즈가 발생하는 입력영상에 대해 Optical flow에 의한 지역 움직임(local motion) 분석기법과 신경회로망을 이용한 상관 추적기법을 적용한 결과, 다중 마커를 정확하고 빠르게 추적하여 얼굴용 광학방식 모션캡처시스템에 적용 가능하다고 본다.

실제 상용화된 외국 타사 제품의 경우 240 ~ 1000 frame/sec 정도까지 처리할 수 있는 고속의 카메라와 특수 전용 보드클 이용하고 있다. 본 연구에서는 외국 타사 제품의 경우처럼 고속의 카메라에 의존하는 방식을 벗어나 일반 저가의 카메라를 이용한 얼굴용 모션캡처시스템을 개발하고자 한다. 따라서 본 연구의 결과로 볼 때 최소 초당 30 프레임 내외를 처리할 수 있는 프레임 그레버의 성능향상과 얼굴의 각 부분에 대한 특징을 고려한 추적 기법이 제시되어야 될 것이다.

참고문헌

- [1] 김용순, 김영수, "3차원 캐릭터 애니메이션 기술 동향," 정보과학회지, 제17권, 제2호, pp.48~59, 02. 1999.
- [2] 이철희, "신경회로망을 이용한 다중 표적의 추출과 추적," 부산대학교 이학석사학위논문, 02. 1999.
- [3] Michael J. Potel, "The Mystery of Motion Capture," IEEE Computer Graphic and Application, pp.14~19, September/October. 1998.
- [4] 서울대학교, "실시간 표적인식 및 추적기법 연구," 자동제어특화연구센터, pp.8~18, 02. 1996.
- [5] 김형재, 이분희, 남양일, 이도엽, 차의영, "신경회로망을 이용한 동적표적의 추적에 관한 연구," 한국정보처리학회 99 춘계학술발표논문집, pp.1355~1358, 04. 1999.
- [6] 이도엽, 이분희, 남양일, 김형재, 차의영, "신경회로망을 이용한 실시간 표적의 추출," 한국정보과학회 99 춘계학술발표논문집, pp.469~471, 04. 1999.
- [7] 박선영, 이상호, 고옥, "실린더 맵핑을 이용한 얼굴 표정의 애니메이션," 한국정보과학회, pp.607~610, 10. 1997.
- [8] 이도엽, 이분희, 김형재, 이철희, 차의영, "ITS용 영상검지기를 위한 다수차량의 실시간 Labeling과 Tracking", 제어·자동화·시스템공학회 제13차 한국자동제어학술회의 논문집, pp.969~972, 10. 1998.
- [9] P.Ekman and W.Friesen, "Facial Acion Coding System," Consulting Psychologists, 1978.

- [10] George Maestri, "Digital Character Animation," Brain&Computer, 1997.
- [11] D. A. Montera, S. K. Rogers, D. W. Ruck, and M. E. Oxley, "Object tracking through adaptive correlation," Optical Engineering, vol. 33, no 1, pp.294~302, 01, 1994.
- [12] 정철근, 김경수, 김중규, "자동영역확장 알고리즘을 이용한 움직이는 객체의 자동 분할", 한국통신학회논문지, Vol.26 No.2B, pp.187~193, 2001.
- [13] B.K.P.Horn B.G.Schunck, "Determining Optical Flow", Artificial Intelligence 17, pp.185~203, 1981.
- [14] A.M.Tekalp, Digital Video Processing, pp.72~94, Prentice Hall, 1995.