

# 실시간 얼굴 특징 점 추출을 위한 색 정보 기반의 영역분할 및 영역 대칭 기법

최승혁<sup>o</sup>, 김재경, 박준\*, 최윤철

연세대학교 컴퓨터과학과,

홍익대학교 컴퓨터공학과\*

{alienart<sup>o</sup>, ki187cm, ycchoy}@rainbow.yonsei.ac.kr, junpark@hongik.ac.kr\*

## Real-Time Face Extraction using Color Information based Region Segment and Symmetry Technique

Sunghyuk Choi<sup>o</sup>, Jaekyung Kim, Jun Park\*, Yoon-Chul Choy

Department of Computer Science, Yonsei University

Department of Computer Engineering, Hong-Ik University\*

### 요 약

최근 가상환경에서 아바타의 활용이 빠르게 증가하면서 아바타 애니메이션에 대한 연구가 활발히 진행되고 있다. 특히 아바타의 사람과 같은 자연스러운 얼굴 애니메이션(Facial Animation)은 사용자에게 아바타가 살아 있는 듯한 느낌(Life-likeness)과 사실감(Believability)을 심어주어 보다 친숙한 인터페이스로 활용될 수 있다. 이러한 얼굴 애니메이션 생성을 위해 얼굴의 특징 점을 추출하는 기법이 끊임없이 이루어져 왔다. 그러나 지금까지의 연구는 실시간으로 사람 얼굴로부터 모션을 생성하고 이를 바로 3D 얼굴 모델에 적용 및 모션 라이브러리를 구축하기 위한 최적화된 알고리즘 개발에 대한 연구가 미흡하였다. 본 논문은 실제 사람 얼굴 모델로부터 실시간으로 특징 점 인식을 통한 애니메이션 적용 및 라이브러리 생성 기법에 대해 제안한다. 제안 기법에서는 빠르고 정확한 특징 점 추출을 위하여 색 정보를 가공하여 얼굴 영역을 추출해내고 이를 영역 분할하여 필요한 특징 점을 추출하였으며, 자연스러운 모션 생성을 위하여 여러 발생 시 대칭점을 이용한 복구 알고리즘을 개발하였다. 본 논문에서는 이와 같은 색 정보 기반의 영역분할 및 영역 대칭 기법을 제시하여 실시간으로 끊임이 없고 자연스러운 얼굴 모션 라이브러리를 생성 및 적용하였다.

### 1. 서 론

최근 영화나 TV 프로그램에는 수많은 3차원 캐릭터들이 등장하고 있다. 이러한 캐릭터들을 얼마나 쉽고 빠르게 조정 할 수 있는가는 중요한 부분이라 할 수 있다. 이를 해결하기 위해 AML(Avatar Markup Language)[1], CML(Character Markup Language)[2] 등 다양한 상위 레벨 언어들이 연구되고 있다. 즉 사용자는 최소한의 스크립트를 작성하여 3D 캐릭터를 조정하는 것이다. 모션을 생성하는 엔진이 스크립트를 해석하고 이를 바탕으로 애니메이션 엔진이 잘게 쪼개진 정적 모션 데이터들을 조합하여 하나의 완성된 행위를 만들어내는 것이다.

이러한 연구들은 모두 하위 레벨의 모션 라이브러리들을 필요로 한다. 이는 크게 몸체 부분과 얼굴 부분으로 나누어 병렬처리되는데 본 논문에서는 얼굴 부분의 모션 라이브러리를 생성하는 것이 그 목적이다. 이러한 모션 라이브러리를 생성하기 위해서 사람이 직접 얼굴 애니메이션을 생성할 수 있지만 이는 전문성이 요구되고 상당한 시간이 요구된다. 이러한 단점을 극복하기 위하여 본 연구에서는 얼굴 모션 캡처 시스템 (Facial Motion Capture System)을 만들어 제시한다. 주변에서 쉽게 구할 수 있는 PC Camera를 이용하여 빠르고 간편하게 모션 라이브러리를 생성함으로써 작업의 용이성을 증대시켰다. 또한 다양하면서 자연스러운 얼굴 애니메이션 데이터를 실시간으로 만들 수 있게 하였다. 뿐만 아니라 이 데이터는 XML 형식으로 저장 되므로 H-Anim 등 표준에 맞는 데이터로 변형이 가능하도록 설계 되었으므로 호환성이 좋다.

### 2. 관련 연구

영역분할이란 영상 정보에서 동질성을 갖는 부분들끼리 특성화시키는 것을 말한다. 이를 위하여 색상 정보를 분석하여 사용하는 RGB, rgb, HSI 등이 그 예이다. 또한 이러한 색상 정보를 영역분할 방법, 에지를 이용하는 방법 등을 이용하여 특징 점들을 찾아낸다.

#### 2.1 색의 가공

RGB 색상 모델은 빨강, 녹색, 파랑의 기본적인 색상 스펙트럼으로 구성되며, 직교 좌표 시스템(Cartesian Coordinate System)에 기초한다. 영상 정보는 RGB 데이터로 입력이 되어 가공되지 않은 색상 정보이다.

$$I = \frac{1}{3}(R + G + B)$$

$$S = 1 - \frac{3}{R + G + B} \min\{R, G, B\}$$

$$H = \cos^{-1} \left[ \frac{\frac{1}{2}[(R - G) + (R - B)]}{\sqrt{(R - G)^2 + (R - B)(G - B)}} \right]$$

[그림 1] HSI 변환식

HSI 색상 모델은 색상, 채도, 명도의 세 가지 색 특징을 표현하는 원뿔형의 모형인데 이는 빛의 강도에 따른 영향을 덜 받기 때문에 피부 검출 시 주로 채택된다. [그림 1]과 같이

RGB을 HSI로 변환할 수 있다. rgb 색상 모델은 RGB 색상 정보를 [그림3]과 같이 색의 비율로 나타냄으로 색의 대조가 크게 나타나는 곳에서 주로 사용된다.

2.2 영역 추출 알고리즘

영역분할(Region Segmentation)을 이용한 방법은 영상에 대한 누적 분포도를 이용하여 적절한 임계치를 결정한 후, 임계치를 이용해서 영상을 분할하는 방법으로 대상 영상과 배경을 분리하는데 사용된다. 에지(Edge)를 이용한 방법은 영상의 각 화소에서 불연속성을 근거로 한 영상분할 방법이다. 영상 비교 방법은 전 프레임의 영상 정보와 현재 프레임의 영상정보를 비교하여 변하지 않는 부분을 배경부분, 변하는 부분을 추출 후보 영역이라고 가정하여 추출하는 방법이다. 클러스터링(Clustering)을 이용한 방법은 영상 내의 화소들의 갖는 특징 값을 하나의 축으로 하여 생성되는 다차원 공간에서 화소들을 여러 클러스터들로 나누고 각 클러스터를 다시 원래의 영상으로 대응시킴으로써 영상을 분할하는 방법이다. 칼라 영상의 정보나 스펙트럼 정보와 같은 다양한 정보를 함께 이용하여 클러스터링 하는 방법이 계속 연구되고 있다[3].

3. 색 정보와 영역 분할 기법

3.1 얼굴 영역 추출 과정

사람의 얼굴은 감정을 표현하거나 대화를 가능하게 하기 때문에 인체에 있어서 가장 중요한 부위라 할 수 있다. 특히 3D 캐릭터에게 행위를 부여시킬 때 얼굴 표정 없이 음성 정보만 나타내 주는 경우 표현 전달 능력이 급격히 떨어지는 것을 알 수 있다. 이를 해결하기 위하여 얼굴에 표정이나 입 움직임 같은 얼굴 애니메이션을 구현해야한다. 이를 통하여 사람들 간의 상호 이해도와 사실감(reality)을 크게 증가시킬 수 있다.

컴퓨터 그래픽스에서 얼굴 애니메이션을 나타내기 위해 근육에 기반을 두어 근육의 움직임에 따라 피부에 미치는 영향을 고려하여 표정을 생성한다. 특정 근육의 특정 위치에 대한 정보를 받아 얼굴 애니메이션을 생성한다.

본 연구는 모션 라이브러리를 생성하는 것이므로 인식률을 높이는 것은 매우 중요한 부분이다. 이 때문에 얼굴의 특정 부위에 푸른색 마커(Marker)<sup>(1)</sup>를 사용하여 특징 점을 추출하였다. 또한 마커는 최대 9개까지 확장할 수 있도록 설계되었다. 하지만 얼굴 애니메이션 중 입의 움직임 즉 말하는 모션에 중점을 두었기 때문에 다음과 같이 입 주변에 4개의 마커를 붙여 사용하였고 눈동자는 자동으로 추출(tracking) 되도록 하였다[4].



a) 필요한 얼굴 영역 b) 마커의 위치 c) 추출된 얼굴 영역 d) 추출 알고리즘 [그림2] 추출 알고리즘

배경에 대한 간섭을 최소화하기 위하여 우선 얼굴 후보 영역을 추출해낸다. 여기서 빛의 영향을 최소화할 수 있는 HSI 색상 모델을 채택 하는 경우가 많지만 본 연구에서는 피부색과 푸른 마커를 구분해야하는 점 그리고 실시간 검출을 위하여 계산식이 간단해야 한다는 점을 고려하여 [그림3]과 같이 정규화 과정을 통하여 색의 비율로 가공하였다. 이를 통하여 휘도 성분에 의한 영향을 배제할 수 있으며 피부에서 마커의 검출도를 높일 수 있다. 그리고 빛에 대한 영향을 적게 하기 위하여 RGB값으로 빛의

<sup>(1)</sup> 얼굴색과 가장 구분되는 색

강도를 측정하고 이에 따라 rgb의 범위를 다르게 주었다. 이를 위해서 RGB와 rgb를 혼용하여 사용하였다.

$$r = R/(R+G+B), g = G/(R+G+B), b = B/(R+G+B)$$

[그림3] RGB 색상의 가공

우선 색의 분포도를 조사하여 rgb별 얼굴 추출 대상 범위를 조사하고 이 범위 내에 들어간 영역을 [그림2] c)와 같이 추출한다. 본 연구에서는 여러 환경 중 일반 가정 즉 형광등 조영에서 최적화 되었고 그 색의 범위는 [그림4]와 같다

색 범위	
피부	0.33 < R < 0.5 & r > (g+10) & g > 50 & b > 50 & g < 240 & b < 240
마커	B > 0.4 & R < 0.25 & b > 29 & r < 230 & g < 230 & g > 29
눈	b < 80 & r < 80 & g < 80

[그림4] 최적화된 색상 구역

얼굴 후보 영역이 검출이 되면 오버 추출, 추출 실패 등의 노이즈들이 나타나게 되는데 이를 제거하기 위하여 영역 성장(Region Growing) 알고리즘을 사용하였다. 이는 소 영역이 서로 간에 같은 특징을 갖고 있는 경우에 그것들을 하나의 영역으로 통합하는 것으로 조금씩 성장시켜 최종적으로 원하는 영역을 얻는 방법이다. 하지만 이는 모든 영역을 탐색해야 하기 때문에 속도가 느려 정지 영상에 적합한 알고리즘이다. 이를 동영상에 적합하게 하기 위하여 [그림2] d)와 같이 대상 영역을 선형으로 검색하게 함으로써 좀 더 효율적이고 빠른 추출이 가능하도록 하였다. 이러한 알고리즘으로 필요 없는 영역에서 오버 추출된 노이즈들을 없애고 추출 실패한 부분을 영역 안으로 넣기 때문에 주위 환경의 영향을 최소화 할 수 있다.

3.2 얼굴의 특징 점 추출 과정

추출된 영역 안에서 특징 점의 의미를 부여하기 위하여 영역을 분할한다. 우선 [그림5] a)와 같이 코를 기준으로 위는 눈동자, 아래는 입으로 규정짓는다. 코 윗부분은 눈 추출에 적합한 색 영역을 주어 눈을 뜨고 있는지 감고 있는지를 감지하고 코 아래는 마커에 대한 색 영역을 주어 입 주변의 특징 점 위치를 감지한다. 이 중 입의 특징 점의 경우 이 그 점이 입의 어느 위치에 있는 점인지 인식할 필요가 있다. 이를 위해서 [그림 5] b)와 같이 코 아래 부분을 영역 분할하여 특징점이 정확히 입의 어느 위치에서 추출되었는지를 분석한다. 이로써 눈의 상태, 그리고 입모양의 상태를 정확하게 추출해 낼 수 있는 것이다.



a) 얼굴 영역 b) 특징 점 추출 [그림5] 특징 점 추출

3.3 Facial Motion Library

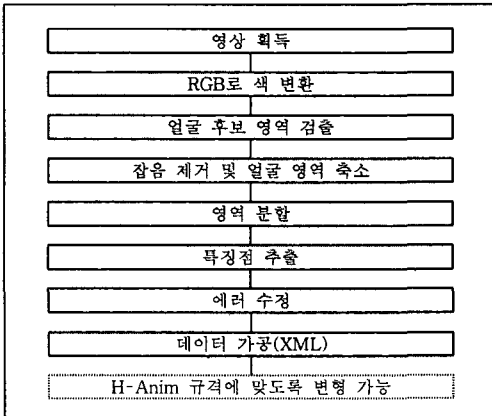
이렇게 추출해낸 특징점도 에러가 발생할 수 있다. 이러한 에러는 캐릭터 애니메이션 대입 시 얼굴의 모양이 일그러지는 현상

을 초래한다. 이를 막기 위하여 여러 복귀 알고리즘이 필요하다. 본 연구에서는 추출 데이터를 분석하여 특징 점 정보가 누락되어 있을 경우 상하 대칭, 좌우 대칭 점을 이용하여 복원시키는 알고리즘을 사용한다. 또한 대칭점 모두가 누락될 경우 전 프레임의 값을 넣어줌으로써 어색한 부분을 막을 수 있다.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE Face SYSTEM "FacialMotion.dtd">
<Face>
  <Motion Level="1" Range="0">
    <Rmouth x="37" y="45"/>
    <Lmouth x="1-35" y="43"/>
    <Umouth x="3" y="13"/>
    <Dmouth x="2" y="73"/>
  </Motion>
  <Motion Level="1" Range="0">
    <Rmouth x="36" y="42"/>
    <Lmouth x="1-35" y="41"/>
    <Umouth x="4" y="15"/>
    <Dmouth x="1" y="71"/>
  </Motion>
</Face>
```

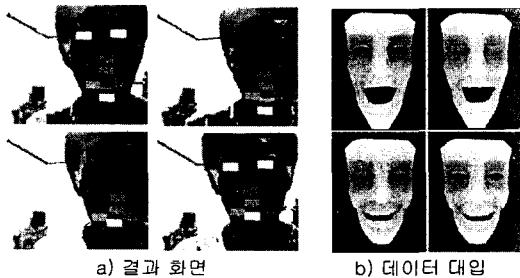
[그림6] XML 저장 형식

위에서 추출된 정보를 [그림6]과 같이 XML[5] 파일로 저장하도록 하였다. XML 파일로 저장이 되기 때문에 XSLT[6]등을 이용하여 데이터의 가공이 용이하고 다른 형식의 문서로 변환이 가능하므로 호환성을 극대화 할 수 있다. [그림7]과 같이 데이터 가공 후 H-anim[7]의 얼굴 표준 규격에 맞도록 변형이 가능하다.



[그림7] Facial Motion Capture System 흐름도

3.4. 구현 결과



[그림8] 결과 화면

[그림8] a)는 위에서 언급한 알고리즘을 도입하여 나타난 결

과화면이다. 특징 점 추출은 96%의 인식률을 보였다. 또한 특징 점 추출에서 있는 인식되지 못한 부분을 여러 수정 과정을 통하여 보완할 수 있다. 이 때문에 3D 모델 적용 시 입 모양이 튀거나 뒤틀리는 현상을 방지할 수 있다.

3D 캐릭터에 데이터를 대입시킬 경우 XML로 저장된 추출 데이터를 3차원 공간에 적합한 데이터로 변형시키고 이를 3D 모델에 적용한다. 이때 얼굴 근육을 기반으로 피부 영역을 나누고 각 영역별로 점들을 그룹화 한다[8]. 얼굴의 특징 점을 중심으로 좌표를 이동시키고 그룹별로 값을 보간(interpolation)하여 자연스러운 애니메이션을 생성한다. 이러한 방법으로 [그림8] b)와 같은 자연스러운 얼굴 애니메이션 결과를 얻을 수 있었다.

4. 결론 및 향후 과제

본 논문에서는 얼굴 모션 라이브러리를 생성하기 위하여 다음과 같은 방향을 제시하였다.

첫째, 과거의 연구는 정지 영상에서 정확한 얼굴 영역을 찾아 내는 것에 초점이 맞추어 있었기 때문에 동영상처럼 실시간 처리를 요하는 것에는 부적합하였다. 이를 위하여 빠르고 효율적인 선영 영역 성장 알고리즘을 이용하였다. 둘째, 정확한 특징 점 추출을 위하여 영역 분할을 하여 주변 환경에 최대한 영향을 덜 받도록 한다. 셋째, 모션라이브러리로 사용하기 위하여 특징 점 추출에 실패한 부분에 대하여 여러 처리를 함으로써 모션이 좀 더 자연스럽다. 넷째, 저장 방식을 XML을 채택함으로써 데이터의 가공 및 호환성이 좋다.

향후 연구 과제로는 모션 라이브러리를 DB화하여 이를 기반으로 필요한 모션 데이터 추출하고 합성하여 Conversation Avatar[9]로 발전시키는 연구이다. 주어진 상황에 따라 아바타가 자동으로 말하는 애니메이션이 생성될 뿐만 아니라 음성과 움직임이 동기화(synchronization)되는 좀 더 사람에 근접한 지능형 아바타 연구가 필요하다.

참고문헌

- [1] Sumedha Kshirsagar, Anthony Guye-Villeme, Kaveh Kamyab Ebrahim Mamdani, Avatar Markup Language, Eight Eurographics Workshop on Virtual Environment, 2002
- [2] Yasmine Arafa, Kaveh Kamyab Ebrahim Mamdani, Sumedha Kshirsagar Nadia Magnenat-Thalmann, Anthony Guye-Vuillème Daniel Thalmann, Two approaches to Scripting Character Animation
- [3] 황윤, 최철, 히스토그램과 영역분할 기법을 이용한 얼굴추출에 관한 연구, 한국정보처리학회 추계 논문집 제9권, 제2호, 2002.11
- [4] Bon-Woo Hwang, Jeong-Seon Park, Seong\_Whan Lee, A Face Synthesis Model Using a Small Number of Feature Points,
- [5] <http://w3c.org/XML/>
- [6] <http://w3c.org/Style/XSL/>
- [7] <http://www.h-anim.org/>
- [8] Jun-young Noh, Ulrich Neumann, Expression Cloning, SIGGRAPH, 2002
- [9] Matthew Stone, Doug DeCarlo, Insuk Oh, Christian Rodriguez, Adrian Stere, Alyssa Lees, Chris Bregler, Speaking with Hands: Creating Animated Conversational Characters from Recordings of Human Performance, SIGGRAPH, 2004