

주색상과 특징점을 이용한 애니메이션 캐릭터의 표정인식

장 석 우[†] · 김 계 영^{**} · 나 현 숙^{***}

요 약

본 논문에서는 사람의 표정이 아닌 애니메이션 캐릭터의 표정을 주색상과 특징점을 효과적으로 분석하여 인식하는 방법을 제안한다. 제안된 방법에서는 먼저 캐릭터의 특성에 맞게 간략화된 메쉬모델을 정의하고 캐릭터 얼굴과 얼굴의 구성요소를 주색상을 이용하여 검출한 후 각 구성요소의 에지를 활용하여 표정인식을 위한 특징점을 추출한다. 그런 다음, 각 특징점의 위치와 모양 정보를 신경망 학습을 통해 해당 AU로 분류하고, 제안된 표정 AU 명세서를 이용해 최종적으로 표정을 인식한다. 실험에서는 제안된 애니메이션 캐릭터의 표정인식 방법이 무표정을 포함하여 기쁨, 슬픔, 놀람, 화남, 공포의 6가지 표정을 비교적 신뢰성 있게 인식함을 애니메이션 영상을 이용한 실험을 통해 보여준다.

키워드 : 애니메이션 캐릭터, 얼굴 표정, 특징점, 신경망

Recognition of Facial Expressions of Animation Characters Using Dominant Colors and Feature Points

Seok-Woo Jang[†] · Gye-Young Kim^{**} · Hyun-Suk Na^{***}

ABSTRACT

This paper suggests a method to recognize facial expressions of animation characters by means of dominant colors and feature points. The proposed method defines a simplified mesh model adequate for the animation character and detects its face and facial components by using dominant colors. It also extracts edge-based feature points for each facial component. It then classifies the feature points into corresponding AUs(action units) through neural network, and finally recognizes character facial expressions with the suggested AU specification. Experimental results show that the suggested method can recognize facial expressions of animation characters reliably.

Keywords : Animation Character, Facial Expression, Feature Point, Neural Network

1. 서 론

최근 들어 애니메이션 산업이 문화 산업과 연계되어 발전하면서 다양한 연령층을 대상으로 한 많은 캐릭터들이 등장하고 있다. 일반적으로, 애니메이션 캐릭터의 감정(emotion)과 성격(personality)을 가장 잘 표현할 수 있는 부분 중의 하나는 캐릭터의 얼굴이다. 따라서 캐릭터의 얼굴과 얼굴 구성요소를 효과적으로 분석하고 활용하는 방법이 필요한데, 특히 이런 방법 중에서 캐릭터 얼굴의 표정을 분석하고 인식하는 기법은 기 제작된 또는 향후 제작될 애니메이션 캐릭터의 얼굴을 사용자의 얼굴로 대체하여 재생해 주는 사용자 맞춤형 애니메이션의 자동 생성 분야에서 그 활용도가

매우 높을 것으로 기대된다.

현재까지 애니메이션 캐릭터 얼굴의 표정을 자동으로 분석하는 기존의 연구는 찾아보기 어렵고[1], 반면에 사람의 얼굴 영상을 받아들여 표정을 분석하고 인식하는 연구는 관련문헌에 소개되어 있다. [2,3]에서는 FACS(facial action coding system)로 얼굴표정에 대한 체계적인 기술 절차를 제시하였으며, 46개의 AU(action unit)를 사용하여 얼굴의 표정을 묘사하였다. [4]에서는 PCA와 FLD(fisher's linear discriminant) 등을 이용하여 얼굴의 특징점을 추출하고, 템플릿 정합을 사용해 6개의 상단얼굴 AU와 6개의 하단얼굴 AU를 구분하였다. [5]에서는 사각 특징점(rectangle feature)을 이용하고 Adaboost 방법을 분류기로 사용하여 얼굴의 표정을 분류하였다.

이와 같이 기존의 얼굴영역이나 표정을 분석하는 많은 방법들은 대부분 사람을 대상으로 하고 있으며, 이 방법들을 애니메이션 캐릭터에 그대로 적용하기에는 많은 무리가 있다. 따라서 본 논문에서는 기존의 얼굴 메쉬모델의 형태를

† 정 회 원 : 안양대학교 디지털미디어학과 조교수
** 중 신 회 원 : 숭실대학교 IT대학 컴퓨터학부 부교수
*** 정 회 원 : 숭실대학교 IT대학 컴퓨터학부 부교수(교신저자)
논문접수: 2011년 5월 23일
수 정 일: 1차 2011년 8월 8일, 2차 2011년 9월 6일
심사완료: 2011년 9월 6일

캐릭터의 특성에 맞게 간략화한 캐릭터 메쉬모형을 정의하고, 주요 색상 정보를 이용하여 얼굴의 특징점을 추출하며, 추출된 특징점들의 위치와 형태 정보를 신경망을 이용하여 학습함으로써 캐릭터의 표정을 인식하는 방법을 새롭게 제안한다. (그림 1)은 제안된 애니메이션 캐릭터 얼굴의 표정 인식 방법의 전체적인 개요도를 보여준다.

(그림 1)에서 보는 바와 같이 제안된 시스템은 크게 표정 학습 단계와 표정인식 단계로 구성된다. 먼저, 표정학습 단계에서는 다양한 색상, 크기 및 형태를 가지는 캐릭터의 얼굴을 강건하게 검출하기 위해서 제안된 캐릭터 얼굴 메쉬모형을 이용해 캐릭터의 얼굴과 얼굴 구성요소의 주색상을 추출하고, 각 구성요소별로 특징점의 위치를 탐색한다. 그리고 특징점의 위치와 형태를 분석하여 22개의 AU를 신경망을 기반으로 학습한다. 본 논문에서 사용하는 캐릭터의 얼굴과 얼굴 구성요소의 주색상을 얻기 위한 메쉬모형은 향후 표정인식을 위한 특징모델로 사용되며, 표정인식에 필요한 눈썹, 눈, 입의 특징점만으로 구성된 메쉬모형이다.

표정인식 단계에서는 학습 단계에서 추출된 주색상을 이용하여 특정 캐릭터의 얼굴과 얼굴 구성요소의 후보영역을 검출한다. 또한, 주색상 값의 유사도 비교를 통해 검출된 얼굴

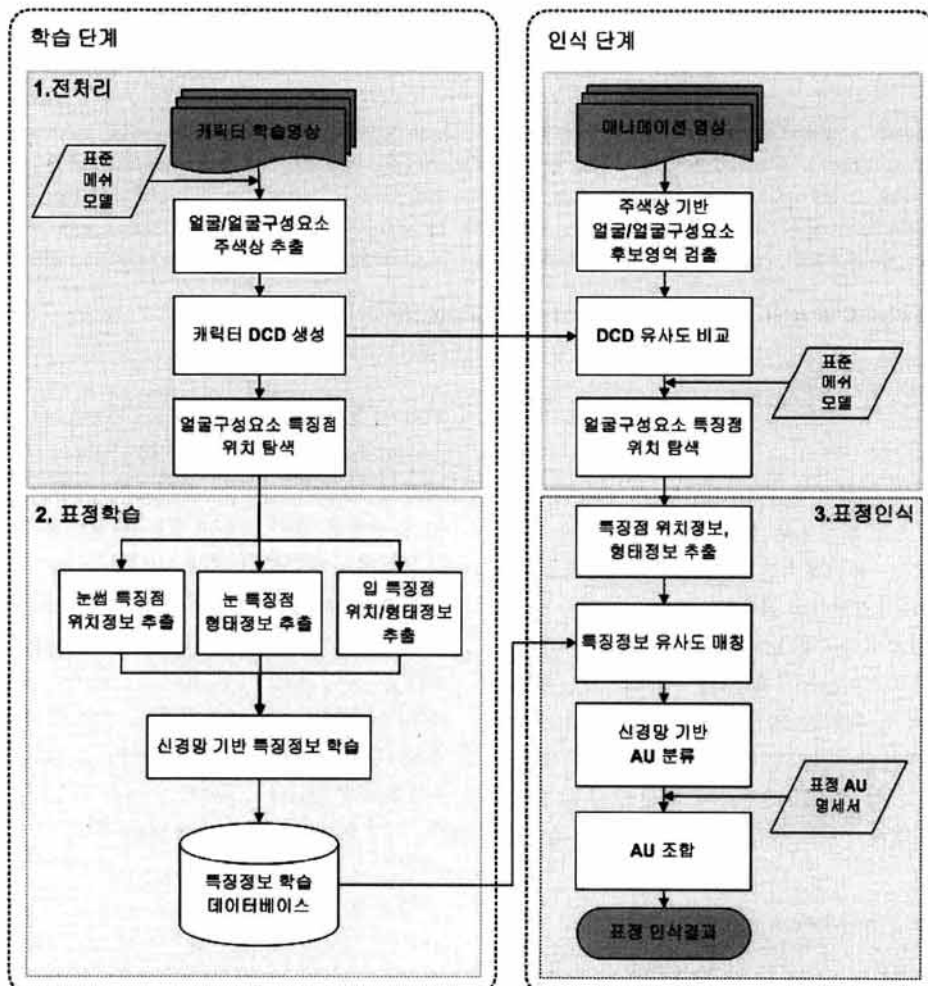
후보영역들 중에서 찾고자 하는 캐릭터 얼굴영역만을 검출한다. 검출된 캐릭터의 얼굴 구성요소 영역에서 에지검출 후 수평과 수직 프로젝션 과정을 수행하여 표정인식에 필수요소인 특징점의 위치를 탐색한다. 마지막으로, 탐색된 특징점들의 위치 및 형태정보를 신경망에 의해 22개의 AU로 분류하고, 표정 AU명세서를 이용하여 무표정을 포함하여 기쁨, 슬픔, 놀람, 화남, 공포의 6가지 캐릭터의 얼굴표정을 인식한다.

본 논문은 총 5장으로 구성된다. 1장에서는 본 연구를 수행하게 된 동기 및 배경, 그리고 전체적인 개요를 기술하였다. 2장에서는 주색상을 이용하여 애니메이션 캐릭터의 특징점을 추출하는 방법을 소개하고, 3장에서는 추출된 특징점을 분석하여 캐릭터의 표정을 인식하는 알고리즘을 설명한다. 그리고 4장에서는 본 논문에서 수행한 실험 결과를 보이며, 5장에서는 결론 및 향후연구 방향을 제시한다.

2. 얼굴 특징점 추출

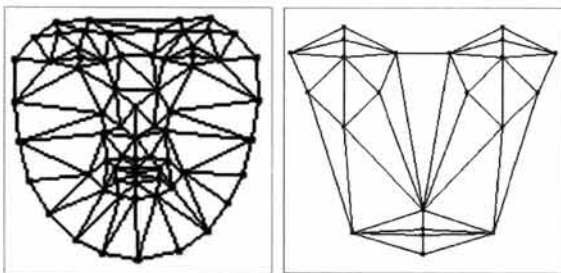
2.1 얼굴과 얼굴 구성요소 검출

애니메이션 캐릭터의 표정을 인식하기 위해서는 얼굴 및 얼굴 구성요소, 그리고 특징점 검출이 선행되어야 한다. 먼



(그림 1) 전체 개요도

저, 본 논문에서는 (그림 2(a))의 기존 얼굴 메쉬모델의 형태를 애니메이션 캐릭터의 특성에 맞게 간략화한 캐릭터 얼굴 메쉬모델을 (그림 2(b))와 같이 정의한다. 이 메쉬모델은 총 22개의 특징점으로 구성되는데, 눈썹 영역에 8개의 특징점, 눈 영역에 8개의 특징점, 입 영역에 6개의 특징점을 가진다. 기존의 메쉬모델은 사람의 형태를 학습하고 인식하기 위한 모델로 많은 특징정보를 추출할 수 있도록 특징점의 개수가 많고, 그 구조 또한 복잡하였다. 그러나 애니메이션 캐릭터의 얼굴은 사람의 얼굴과는 달리 얼굴 구성요소의 형태가 일정한 비율을 가지지 못하고 매우 다양하므로 모든 캐릭터에 대해 표준적인 얼굴모델을 생성하기가 어렵다. 또한, 얼굴 구성요소의 형태가 단순하고, 색상이나 예지 정보 이외에 텍스처와 같은 복잡한 특징을 포함하지 않기 때문에 기존의 얼굴 메쉬모델을 그대로 이용하기 보다는 캐릭터 얼굴에 특성화된 메쉬모델이 필요하다. 따라서 본 논문에서는 애니메이션 캐릭터 얼굴의 3가지 구성요소(눈썹, 눈, 입)를 부각시키고, 나머지 구성요소는 제거함으로써 필수적인 특징점들로만 구성된 애니메이션 캐릭터의 메쉬모델을 생성한다.



(a) 기존의 메쉬모델 (b) 제안된 메쉬모델
(그림 2) 메쉬모델

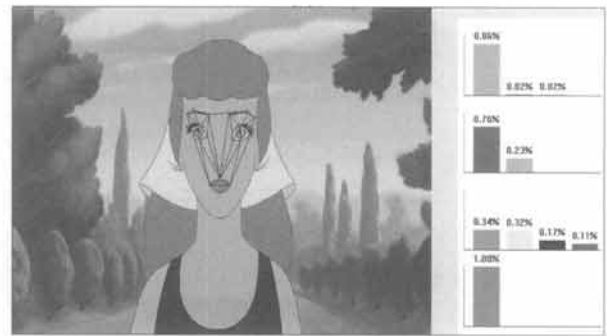
그런 다음, 캐릭터의 얼굴을 검출하기 위해 캐릭터가 가진 고유한 특징 중의 하나인 색상정보를 활용하는데, 색상 값으로는 MPEG-7에 정의된 주색상 서술자인 DCD (dominant color descriptor)를 사용한다[6,7]. 보통, 주색상 서술자 F 는 식 (1)과 같이 각 주색상의 값 c_i , 주색상 값을 가지는 화소의 비율 p_i , 주색상 값의 분산인 v_i , 그리고 공간적인 색상의 응집도인 s 로 정의된다. 그리고, i 는 주색상을 나타내는 인덱스이다.

$$F = \{(c_i, p_i, v_i), s\}, \quad (i = 1, 2, \dots, N) \quad (1)$$

본 논문에서는 얼굴 영역의 주색상을 추출하기 위해 그림 3과 같이 입력된 원본영상으로부터 애니메이션 캐릭터의 얼굴 구성요소에 맞도록 제안된 메쉬모델을 수작업으로 정합시킨다. 그런 다음, 메쉬모델에 포함된 얼굴 구성요소 영역을 제외한 메쉬모델 영역에서 DCD를 사용해 얼굴의 주색상을 추출한다. 일반적으로, 이 서술자는 임의의 영역에 대한 주요색의 집합을 나타내며, 최소 1개에서 최대 8개 사이의 주색상을 나타낸다. 본 논문에서는 애니메이션 캐릭터 얼굴

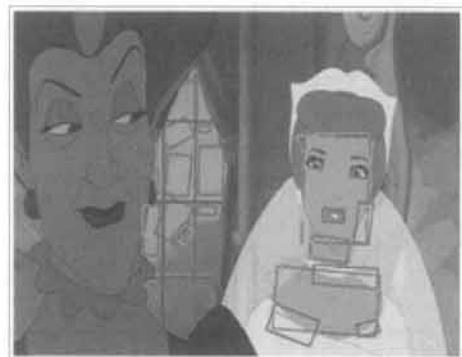
영역의 주색상 추출 시 영역 내에서 차지하는 비율이 10% 미만인 색상은 제외하였다. 그리고 추출된 캐릭터 얼굴은 보통 한 가지의 주색상으로 구성된다.

캐릭터의 얼굴 구성요소는 얼굴과 달리 하나의 주색상이 아닌 몇 가지 주색상의 조합으로 구성된다. 얼굴 구성요소의 주색상은 제안된 메쉬모델이 포함하고 있는 각 구성요소 영역에서 유사한 색상 값을 가지는 영역을 병합하는 과정을 통해 각 구성요소의 주색상을 추출한다. (그림 3)에서 확인할 수 있듯이, 얼굴 영역과 입술 영역은 하나의 주색상, 눈썹은 2개의 주색상, 그리고 눈은 4개의 주색상이 추출되었다. 이와 같이 추출한 주색상을 이용하여 향후 새롭게 입력되는 영상으로부터 얼굴과 얼굴 구성요소의 후보영역을 검출할 수 있다.



(그림 3) 주색상 추출

이와 같이 추출한 주색상을 이용하여 새롭게 입력되는 영상으로부터 얼굴과 얼굴 구성요소의 후보영역을 (그림 4)와 같이 검출하고, 검출된 후보 영역들 각각의 최소 경계 사각형(MBR: minimum bounding rectangle)[8]을 구한다. 이때 얼굴 후보영역이라고 판단된 MBR은 얼굴 구성요소에 대한 MBR을 모두 포함하고 있어야 한다.



(그림 4) 얼굴 구성요소 후보영역 추출

그런 다음, 검출된 얼굴 후보 영역 중에서 실제 캐릭터 얼굴 영역을 찾기 위해서 알고리즘 초기에 설정한 얼굴 구성요소의 주색상과 후보 구성요소 영역들의 주색상과의 유사도를 식 (2)를 이용하여 비교한다. 식 (2)에서 F_1 과 F_2 는 두 개의 DCD 기술자를 나타내고 있고, N_1 과 N_2 는 F_1 과

F_2 에서 사용하는 색의 개수, 그리고 p_{1i} 와 p_{2i} 는 F_1 과 F_2 의 i 번째 주색상의 비율을 나타낸다. $c_{xi}^l, c_{xi}^u, c_{xi}^v$ 는 x 번째 기술자, i 번째 주색상의 L값, U값, V값[6]을 나타내고, $v_{xi}^l, v_{xi}^u, v_{xi}^v$ 는 x 번째 기술자, i 번째 주색상의 L값, U값, V값의 분산을 나타낸다.

$$D_v(F_{1,F_2}) = \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{1i} p_{1j} f_{1i,1j} + \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} p_{2i} p_{2j} f_{2i,2j} - \sum_{i=1}^{N_1} \sum_{j=1}^{N_2} 2p_{1i} p_{2j} f_{1i,2j}$$

$$f_{\xi,yj} = \frac{1}{2\pi \sqrt{v_{\xi,yj}^l v_{\xi,yj}^u v_{\xi,yj}^v}} \cdot \exp \left[- \left(\frac{c_{\xi,yj}^l}{v_{\xi,yj}^l} + \frac{c_{\xi,yj}^u}{v_{\xi,yj}^u} + \frac{c_{\xi,yj}^v}{v_{\xi,yj}^v} \right) / 2 \right]$$

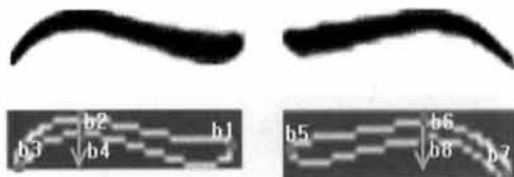
$$c_{\xi,yj}^l = (c_{\xi}^l - c_{yj}^l)^2, c_{\xi,yj}^u = (c_{\xi}^u - c_{yj}^u)^2, c_{\xi,yj}^v = (c_{\xi}^v - c_{yj}^v)^2$$

$$v_{\xi,yj}^l = (v_{\xi}^l - c_{yj}^l), v_{\xi,yj}^u = (v_{\xi}^u - c_{yj}^u), v_{\xi,yj}^v = (v_{\xi}^v - c_{yj}^v) \quad (2)$$

본 논문에서는 식 (2)를 이용하여 오차범위 $\pm 10\%$ 내에서 유사도 매칭이 수행된다. 만약, 영상 내에서 검출된 후보영역들이 얼굴 구성요소의 주색상 정보를 모두 포함하는 조건을 만족하지 못한다면 그 영역은 제외되며, 최종적으로 가장 유사한 값을 가지는 영역만이 찾고자 하는 애니메이션 캐릭터의 얼굴 영역이라고 판단된다.

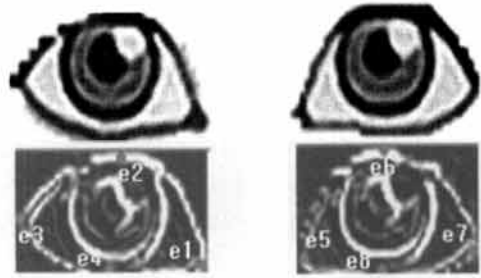
2.2 얼굴 특징점 추출

보통 캐릭터의 얼굴은 사람과 달리 에지의 성분이 뚜렷하고, 조명이나 잡음의 영향을 상대적으로 적게 받는다는 특징이 있다. 따라서 본 논문에서는 추출된 얼굴 구성요소 내에서 수평과 수직 에지에 민감하게 반응하는 프리윗(Prewit) 에지[9] 검출기를 이용하여 특징점의 위치를 탐색한다. 먼저, 눈썹 영역의 특징점을 추출하기 위해 에지 검출 및 수평 프로젝션을 수행하여 눈썹의 좌우 양 끝점을 특징점으로 결정하며, 수직 프로젝션을 수행하여 상단의 특징점을 결정한다. 마지막 특징점은 상단 특징점의 하향 수직으로 가장 가까운 에지를 특징점으로 결정한다. (그림 5)는 이와 같이 찾아진 눈썹 특징점의 예를 보여준다.



(그림 5) 눈썹 특징점(8개)

눈의 특징점은 눈썹의 특징점과 유사하게 좌우 양 끝점과 상단 및 하단의 에지를 (그림 6)과 같이 특징점으로 추출한다. 눈의 경우 눈썹이나 입과 달리 그 형태나 회전의 변화가 심하지 않고 주로 눈의 크기나 끝점의 이동으로 형태가



(그림 6) 눈 특징점(8개)

결정되기 때문에 좌우상하의 특징점만으로도 표정분석이 가능하다.

입은 (그림 7)과 같이 총 6개의 특징점을 결정하기 위하여 입 영역 내에서 에지를 검출한다. 그리고 수평 프로젝션을 수행하여 양단의 끝점을 입 꼬리의 특징점으로 결정한다. 입의 모양은 윗입술에서의 두 개의 입술산으로 인해 상단의 특징점을 결정하기 어려우므로 하단의 특징점을 기반으로 MBR 영역 내에서 상향 수직으로 에지를 탐색한다. 이는 입의 형태가 사람과 달리 입술산이 존재하거나 존재하지 않는 캐릭터에도 적용이 가능한 방법으로 회전되거나 입술선이 뚜렷하게 그려지지 않은 경우에도 가능하다. 탐색된 에지들 중 하단에서부터 처음 탐색된 에지와 가장 마지막에 탐색된 에지, 그리고 마지막 에지로부터 바로 이전에 탐색된 에지를 특징점으로 결정한다.



(그림 7) 입 특징점 (6개)

3. 캐릭터의 표정 인식

캐릭터의 얼굴표정을 인식하기 위해서 이전 단계에서 추출된 특징점들의 위치와 모양을 분석한 후 FACS에서 정의한 AU를 적용한다. 본 논문에서는 캐릭터의 얼굴 형태를 고려하여 구성한 14개의 단일 AU와 8개의 복합 AU, 즉 총 22개의 AU를 신경망을 기반으로 학습하고, 본 논문에서 정의한 표정 AU 명세서를 이용한 AU 조합을 통해 표정을 인식한다.

3.1 AU 정의

(그림 8)은 캐릭터의 표정인식을 위해 본 논문에서 사용하는 얼굴 구성요소들의 단일 AU이다. (그림 8)에서 확인할 수 있듯이 눈썹, 눈, 입에 대한 11개의 AU와 각 구성요소의 무표정을 나타내는 3개의 AU로 구성된다. 그리고 단일 AU를 기반으로 하단의 얼굴에서 나타날 수 있는 복합 AU 8개를 정의하여 사용한다. 본 논문에서는 FACS의 얼굴표정을 묘사하는 46개의 AU 중에서 캐릭터의 얼굴에는 존재하지

| | | | |
|-------------|--------------|--------------|------------|
| Neutral(NB) | AUI | | AU2 |
| | | | |
| 기본적인 눈썹 | 눈썹 내측이 올라감 | | 눈썹 외측이 올라감 |
| Neutral(NE) | AU5 | AU7 | AU43 |
| | | | |
| 기본적인 눈 | 눈을 크게 뜸 | 눈을 살짝 감음 | 눈을 완전히 감음 |
| Neutral(NL) | AU12 | AU15 | AU18 |
| | | | |
| | 입술 양단을 끌어 올림 | 입술 양단을 끌어 내림 | 입술을 오므림 |
| | AU20 | AU25 | AU27 |
| | | | |
| 기본적인 입 | 입술 양단을 늘림 | 입을 작게 벌림 | 입을 크게 벌림 |

(그림 8) Action Units

않는 주름에 관한 AU와 표정분석에 적은 영향을 미치는 코에 대한 AU는 제외하였다. 그리고 <표 1>은 AU로 정의한 3가지 구성요소들의 기본 템플릿을 정의한 것이다. 눈썹의 상태는 존재의 '유/무'로 판단하며, 눈의 상태는 '뜬눈/감은눈'으로, 입은 '열었다/닫았다'의 상태로 표현한다. 템플릿 상의 각 제어점은 AU 제어점이라 정의하고[10], 이것은 표정을 묘사하는 AU를 정의할 때 파라미터로 사용된다.

<표 1> 얼굴 구성요소 템플릿

| 구성요소 | 상태 | 구성요소 템플릿 |
|------|---------|----------|
| 눈썹 | 유/무 | |
| 눈 | 뜨다 / 감다 | |
| 입 | 열다 / 닫다 | |

3.2 AU를 이용한 표정 학습 및 인식

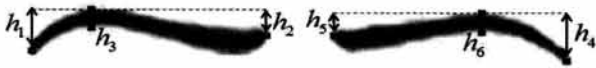
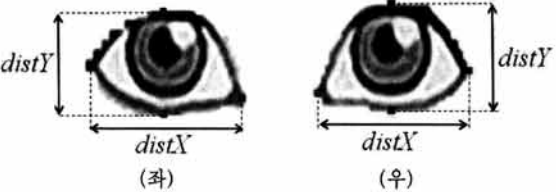
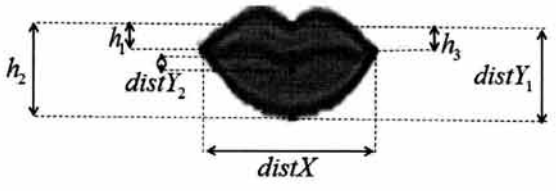
본 논문에서는 22개의 특징점을 각 얼굴 구성요소별 특징점의 가변적인 위치와 모양 정보를 기반으로 표정을 구분한

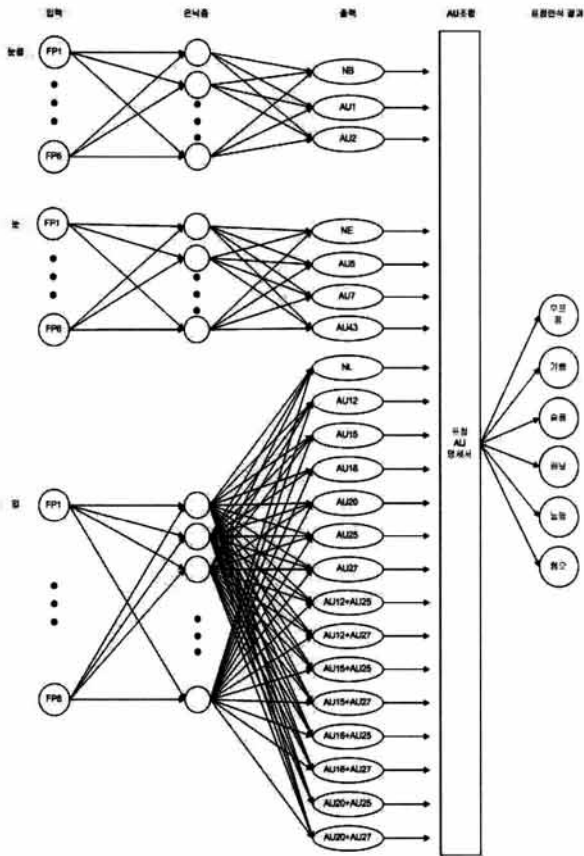
다. <표 2>는 얼굴 구성요소별로 표정학습을 위한 특징을 나타낸다. 제안된 방법에서는 무표정을 정의하는 NB, NE, NL 3개의 AU를 포함한 단일 AU 14개와 하단 얼굴에서 나타나는 복합 AU 8개를 포함한 총 22개의 AU를 학습한다.

각 구성요소들의 특징점 위치정보를 학습하기 위해 (그림 9)와 같은 신경망을 이용한다. 신경망의 입력으로는 <표 2>에서 정의된 각 AU를 구성하는 특징점의 위치와 모양정보가 이용되며, 은닉층 노드의 개수는 눈썹은 16개, 눈은 16개, 입은 21개이다. 그리고 출력 노드는 눈썹 3개, 눈 4개, 입 15개의 얼굴 구성요소별 표정인식을 수행하기 위한 22개의 AU를 출력 값으로 가진다. (그림 9)에서 FP(feature points)는 입력노드를 구분하기 위한 기호이다.

애니메이션 캐릭터 얼굴영상을 수집한 데이터베이스를 이용하여 학습한 표정 AU를 분석한 결과 3가지 얼굴 구성요소 중 눈썹은 특정한 AU의 강도가 높게 나타나므로 단독 AU로 정의하는데 반해, 눈과 입의 AU는 거의 모든 AU가 표정에서 고른 분포를 가진다는 것을 알 수 있었다. 따라서 본 논문에서는 <표 3>과 같이 6가지 표정을 정의하기 위해 각 구성요소의 AU를 조합함으로써 표현의 범위를 정의한다. 예를 들어, '기쁨' 표정을 표현하기 위한 AU는 눈썹, 눈, 입 각각 1개, 4개, 4개의 AU를 조합한 총 16가지의 AU 조합으로 정의된다. 이것은 1차적으로 임의의 실험자들을 대상으로 시각적으로 판단하여 각 표정의 AU를 분류하고, 2차적으로 얼굴 구성요소들의 모양 구조를 분석하여 분류된 결과를 비교함으로써 공통적으로 추출된 AU의 조합으로 정의한 것이다. 이와 같은 방법으로 총 6가지 표정에 대한 AU의 조합을 정의한다. 이와 같이, 출력된 얼굴 구성요소별 AU를 <표 3>에서 제안하는 표정 AU 명세서를 이용하여 조합한 후, 최종적으로 무표정과 5가지의 표정을 인식한다.

〈표 2〉 AU 표정학습

| 구성요소 | AU | 학습방법 |
|------|---------|--|
| 눈썹 | NB | 1. 눈썹 4개(b_1, b_2, b_3, b_4) 포인트들 간의 거리 $h_i = b_j$ 의 y 좌표 - b_2 의 y 좌표 |
| | AU1 | b_2 특징점과 b_1 특징점의 y 좌표의 거리 : h_1 (좌), h_4 (우) b_2 특징점과 b_3 특징점의 y 좌표의 거리 : h_2 (좌), h_5 (우) b_2 특징점과 b_4 특징점의 y 좌표의 거리 : h_3 (좌), h_6 (우) |
| | AU2 | 2. h_1, \dots, h_6 의 거리를 이용하여 눈썹에 대한 AU 3가지에 대해 학습  |
| 눈 | NE | 1. 눈 4개(e_1, e_2, e_3, e_4) 특징점 간의 거리와 비율 e_1 특징점과 e_3 특징점간의 거리 : $distX$ (좌), $distX$ (우) e_2 특징점과 e_4 특징점간의 거리 : $distY$ (좌), $distY$ (우) $distY$ 와 $distX$ 의 비율 : R_{Eye} (좌), R_{Eye} (우) |
| | AU5 | 2. $distX$ (좌), $distX$ (우), $distY$ (좌), $distY$ (우), R_{Eye} (좌), R_{Eye} (우)를 이용해 눈에 대한 AU 4개를 학습 |
| | AU7 |  |
| | AU43 | |
| 입 | NL | 1. 입 6개의($l_1, l_2, l_3, l_4, l_5, l_6$) 특징점 간의 거리와 비율 |
| | AU12 | $h_i = l_j$ 의 y 좌표 - l_2 의 y 좌표 |
| | AU15 | l_2 특징점과 l_1 특징점간 y 좌표의 거리 : h_1 |
| | AU18 | l_2 특징점과 l_5 특징점간 y 좌표의 거리 : h_2 |
| | AU20 | l_2 특징점과 l_6 특징점간 y 좌표의 거리 : h_3 |
| | AU25 | l_1 특징점과 l_6 특징점간의 거리 : $distX$ |
| | AU27 | l_2 특징점과 l_5 특징점간의 거리 : $distY_1$ l_3 특징점과 l_4 특징점간의 거리 : $distY_2$ |
| | AU12+25 | $distX$ 와 $distY_1$ 의 비율 : R_{Lip1} |
| | AU12+27 | $distX$ 와 $distY_2$ 의 비율 : R_{Lip2} |
| | AU15+25 | |
| | AU15+27 | 2. h_1, h_2, h_3 의 거리차와 $distX, distY_1, distY_2, R_{Lip1}, R_{Lip2}$ 를 이용하여 눈에 대한 AU 15개를 학습 |
| | AU18+25 |  |
| | AU18+27 | |
| | AU20+25 | |
| | AU20+27 | |



(그림 9) 표정인식을 위한 신경망

<표 3> 제안하는 표정 AU 명세서

| 표정 | 눈 AU | 입 AU |
|-----|------|-----------|
| 무표정 | NB | NL |
| 기쁨 | NB | NE |
| | | AU12 |
| | | AU27 |
| | | AU7 |
| 슬픔 | AU1 | AU43 |
| | | NL |
| | | AU15 |
| | | AU25 |
| 놀람 | AU5 | AU27 |
| | | AU15+AU25 |
| | | AU15+AU27 |
| | | AU18+AU25 |
| 화남 | AU2 | AU18+AU27 |
| | | AU20+AU25 |
| | | AU20+AU27 |
| | | NE |
| | | AU18 |
| | | AU20 |
| | | AU25 |
| | | AU27 |
| AU7 | | |
| 공포 | AU1 | AU15+AU25 |
| | | AU20+AU25 |

4. 실험 결과

본 논문에서는 실험을 위해 인텔 펜티엄 IV 3.0GHz의 CPU와 2GB의 메모리가 장착된 컴퓨터를 사용하였고, 운영 체제는 마이크로소프트사의 Windows XP Professional을 사용하였다. 그리고 소프트웨어 구현을 위한 컴파일러로는 마이크로소프트사의 Visual Studio C++ 6.0과 영상처리 오픈 라이브러리인 OpenCV 1.1 및 OpenGL을 사용하였다. 실험을 위하여 애니메이션의 키 프레임 영상을 이용하며, 캐릭터 얼굴의 구성요소가 다른 캐릭터나 주변 그림들로 인하여 폐색되거나 얼굴이 심하게 회전되지 않은 영상을 사용한다. 또한, 알고리즘 초기에 수행하는 캐릭터의 주색상 정보를 추출하기 위해서 무표정 상태의 캐릭터 영상을 사용한다. 본 논문에서는 75분짜리 신데렐라 애니메이션으로부터 신데렐라 캐릭터의 얼굴을 6가지 표정별로 분류하여 각각 30개씩 추출하고, 각 표정을 대표하는 눈썹, 눈, 입에 대한 특징 값을 추출하여 신경망의 학습 데이터로 사용하였다.

캐릭터의 얼굴 구성요소 영역에서 에지 탐색과 수평 및 수직 프로젝션을 수행하여 각 구성요소의 특징점을 추출한다. 그리고 특징점들의 위치와 형태 정보를 신경망에 입력하여 얼굴 구성요소의 AU를 검출하며, 제안된 표정 AU 명세서를 이용하여 캐릭터의 표정을 최종적으로 인식한다. (그림 10)은 이런 과정을 통해 신데렐라 애니메이션 캐릭터에 대한 무표정과 5가지의 표정을 인식한 결과를 보여준다.



(그림 10) 표정인식 결과

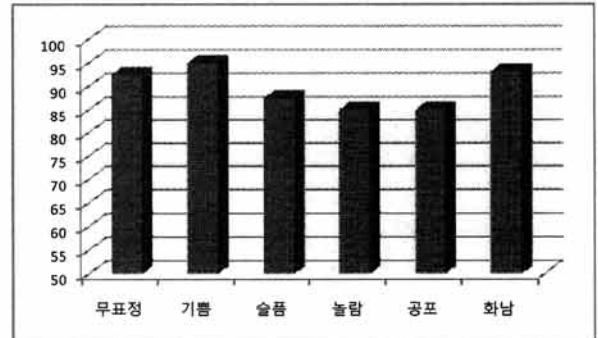
(그림 10)에서 확인할 수 있듯이 캐릭터 얼굴의 회전이나 크기에 강건하게 표정인식이 수행되었으며, 영상 안에 두 개 이상의 캐릭터가 등장하여도 캐릭터마다 고유의 주색상 정보를 가지고 있으므로 대상 캐릭터의 표정인식이 가능하다.

<표 4>는 각 구성요소의 무표정 상태를 구성하는 3개의 AU와 11개의 단독 AU, 8개의 복합 AU에 대해 각각 10개씩 총 220개의 AU에 대해 식 (3)을 이용하여 전체 AU의

인식률과 오 인식률, 그리고 단독 AU와 복합 AU의 인식률과 오 인식률을 측정하였다. 식 (3)에서 $R_{correct}$ 와 $R_{incorrect}$ 는 인식률과 오 인식률을 나타내며, N_{input} 과 $N_{correct}$ 는 입력된 영상의 개수와 표정인식이 정확하게 수행된 영상의 개수를 나타낸다. AU의 평균 인식률은 88%이고 오 인식률은 12%이다. 그리고 단독 AU의 인식률은 93%로 복합 AU의 인식률인 68%보다 높은 인식률을 보였다. 왜냐하면 복합 AU는 주로 입을 묘사하는 AU로서 특징점의 탐색 과정에서 탐색의 오류로 특징점이 잘못 위치할 경우 오 인식률이 높게 나타나기 때문이다.

$$R_{correct} = \frac{N_{correct}}{N_{input}}, \quad R_{incorrect} = 1 - R_{correct} \quad (3)$$

(그림 11)은 애니메이션 캐릭터 표정인식의 성능을 측정하는 것으로 각 표정에 대해 40개씩의 샘플 영상을 사용하였다. 인식의 대상이 되는 캐릭터의 표정 중에서 '공포' 라는 표정



(그림 11) 캐릭터의 표정 인식률

에 대해서 인식률이 전체적으로 낮은 것을 알 수 있는데, 이것은 '놀람' 표정과 '공포' 표정의 경우 캐릭터의 특성 상 주로 눈은 크게 뜨고 입은 벌린 상태가 많아 두 표정이 서로 유사하게 인식되기 때문인 것으로 판단된다. 이를 제외하면 대부분의 표정에서 비교적 높은 표정 인식률을 보인다.

추가적으로 기존의 얼굴인식 기법을 애니메이션 캐릭터에 그대로 적용했을 때와 본 논문에서 제안한 방법을 적용했을

<표 4> AU 인식률

| AU(Action Units) | 샘플(수) | 인식(수) | 오인식(수) | |
|---------------------|-------|-------|-------------------------|-----|
| NB(Neutral eyeBrow) | 10 | 9 | 1(AU1) | |
| NE(Neutral Eye) | 10 | 10 | - | |
| NL(Neutral Lip) | 10 | 9 | 1(AU25) | |
| AU1 | 10 | 10 | - | |
| AU2 | 10 | 10 | - | |
| AU5 | 10 | 9 | 1(NE) | |
| AU7 | 10 | 9 | 1(AU5) | |
| AU12 | 10 | 10 | - | |
| AU15 | 10 | 8 | 2(NL,AU20) | |
| AU18 | 10 | 8 | 2(AU25) | |
| AU20 | 10 | 9 | 1(AU25) | |
| AU25 | 10 | 10 | - | |
| AU27 | 10 | 10 | - | |
| AU43 | 10 | 10 | - | |
| AU12+25 | 10 | 9 | 1(AU12+27) | |
| AU12+27 | 10 | 9 | 1(AU12+25) | |
| AU15+25 | 10 | 8 | 2(AU25,AU15+27) | |
| AU15+27 | 10 | 7 | 3(AU27,AU15+20,AU20+27) | |
| AU18+25 | 10 | 7 | 3(AU25,AU18+AU27) | |
| AU18+27 | 10 | 7 | 3(AU20+25,AU20+27) | |
| AU20+25 | 10 | 8 | 2(AU25) | |
| AU20+27 | 10 | 7 | 3(AU18+AU20,AU20+AU25) | |
| 총 개수 | 220 | 193 | 27 | |
| 인식률(%) | | | | |
| 총 인식률 | | 88% | 총 오인식률 | 12% |
| 단독 AU 인식률 | | 93% | 단독 AU 오인식률 | 7% |
| 복합 AU 인식률 | | 68% | 복합 AU 오인식률 | 22% |

〈표 5〉 기존의 방법과 제안된 방법의 비교

| 방법 | 장점 | 단점 |
|-------------|---|--|
| 기존의 얼굴인식 방법 | - 애니메이션 캐릭터가 실제 사람과 매우 유사한 형태를 가질 때는 정확도 높음 | - 기존의 복잡한 메쉬모델을 사용하므로 계산 복잡도가 높음 - 너무 세분화되어 있어 오히려 정확도가 낮은 부분이 생김 |
| 제안된 방법 | - 단순화된 메쉬모델을 사용하므로 계산 복잡도가 낮음 - 적당히 세분화되어 있으므로 정확도가 대체적으로 좋음 | - 일반적인 피부모델을 적용할 수 없으므로 매 캐릭터마다 서로 다른 주색상을 추출해야 함 |

매의 장점과 단점을 정성적으로 비교 분석하면 <표 5>와 같다.

5. 결론 및 향후연구

본 논문에서는 주색상과 특징점을 효과적으로 분석하여 애니메이션 캐릭터의 표정을 인식하는 방법을 제안하였다. 제안된 방법에서는 메쉬모델을 기반으로 추출한 주색상을 이용해 얼굴과 얼굴의 구성요소를 검출하고, 각 구성요소의 에지를 활용하여 표정인식을 위한 특징점을 추출하였다. 또한, 각 특징점의 위치와 모양 정보를 신경망 학습을 통해 해당 AU로 분류하고, 본 논문에서 제안한 표정 AU 명세서를 이용하여 최종적으로 표정을 인식한다. 기존의 사람 얼굴의 표정을 인식하기 위한 연구들에 비해 캐릭터의 얼굴 특징을 고려하여 복잡한 특징을 이용하는 대신 주색상이라는 캐릭터만의 고유한 정보를 추출하였고, 이를 통하여 간단하고 빠르게 캐릭터의 얼굴 영역을 검출하였다. 또한, 캐릭터마다 얼굴 구성요소의 형태가 다양하다는 특성을 고려하여 학습된 AU를 바탕으로 표정 AU 명세서를 이용하여 단독 AU들을 조합함으로써 복잡적이고 다양한 표정을 인식할 수 있도록 제안하였다.

현재 얼굴 구성요소 특징점의 위치 탐색을 위해 에지검출을 수행하는데, 이때 중요한 에지들이 다른 캐릭터에 의해 가려지는 현상이 발생하거나 미약한 에지들로 인하여 정확한 위치 탐색이 수행되지 못하는 경우가 종종 발생한다. 만약, 입을 벌린 경우 치아로 인한 불필요한 에지들이 잘못 탐색되기도 하며, 윗 입술과 아래 입술의 경계를 정확하게 그리지 않은 캐릭터의 경우 에지 탐색이 어렵다. 향후 연구로는 이를 해결하기 위해 캐릭터의 주색상 정보를 잃지 않고 에지검출을 강화할 수 있는 전처리 단계를 보완할 예정이다. 또한, 인식을 시도할 표정의 종류를 좀 더 유연하게 확장할 수 있는 방안에 대한 연구를 수행하고, 보다 많은 샘플을 이용하여 성능을 평가할 것이다.

참고 문헌

[1] J.-H. Kim, "A Study on the Key Elements for Facial

Expression in Animation Character," In Proceedings of the Spring Conference of the Korea Institute of Design, pp.268-269, 2006.

- [2] P. Ekman, W. V. Friesen, and S. Ancoli, "Facial Signs of Emotional Experience," *Journal of Personality and Social Psychology*, Vol.39, No.6, pp.1125-1134, 1980.
- [3] Y.-I. Tian, T. Kanade, and J. F. Cohn, "Recognizing Action Units for Facial Expression Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol.23, No.2, pp.97-115, 2001.
- [4] G. Donato, M. S. Bartlett, J. C. Hager, P. Ekman, and T. J. Sejnowski, "Classifying Facial Actions," *IEEE Transactions on Multimedia*, Vol.4, No.1, pp.23-37, 2002.
- [5] Y. Wang, H. Ai, B. Wu, and C. Huang, "Real Time Facial Expression Recognition with Adaboost," In Proceedings of the IEEE International Conference on Pattern Recognition, Vol.3, pp.926-929, 2004.
- [6] R. Min and H. D. Cheng, "Effective Image Retrieval Using Dominant Color Descriptor and Fuzzy Support Vector Machine," *Pattern Recognition*, Vol.42, No.1, pp.147-157, 2009.
- [7] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "MPEG-7 Color and Texture Descriptors," *IEEE Transactions on Circuits, Systems and Video Technology*, Vol.11, pp.703-715, 2001.
- [8] M. A. Cobb, F. E. Petry, and K. B. Shaw, "Fuzzy Spatial Relationship Refinements Based on Minimum Bounding Rectangle Variations," *Fuzzy Sets and Systems*, Vol.113, No.1, pp.111-120, 2000.
- [9] Z. Lin, J. Jiang, and Z. Wang, "Edge Detection in the Feature Space," *Image and Vision Computing*, Vol.29, No.2-3, pp.142-154, 2011.
- [10] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features," In Proceedings of the International Conference on Computer Vision and Pattern Recognition, pp.511-518, 2001.



장 석 우

e-mail : swjang@anyang.ac.kr

1995년 숭실대학교 전자계산학과(학사)
1997년 숭실대학교 컴퓨터학과(공학석사)
2000년 숭실대학교 컴퓨터학과(공학박사)
2003년~2008년 한국건설기술연구원
건설정보화연구부 선임연구원

2009년~현 재 안양대학교 디지털미디어학과 조교수
관심분야: 로봇비전, 증강현실, HCI, 비디오 색인 및 검색 등



나 현 숙

e-mail : hsnaa@ssu.ac.kr

1993년 서울대학교 수학과(학사)
1995년 포항공과대학교 수학과(석사)
2002년 포항공과대학교 수학과(박사)
2001년~2002년 프랑스 INRIA Post Doc.
2002년~2003년 홍콩과학기술대(HKUST)
전산학과 Post Doc.

2003년~현 재 숭실대학교 IT대학 컴퓨터학부 부교수
관심분야: 알고리즘, 계산기하학, 정보이론 등



김 계 영

e-mail : gykim11@ssu.ac.kr

1990년 숭실대학교 전자계산학과(공학사)
1992년 숭실대학교 컴퓨터학과(공학석사)
1996년 숭실대학교 컴퓨터학과(공학박사)
1996년~1997년 한국전자통신연구원
Post-Doc.

1997년~2001년 한국전력 선임연구원
2001년~현 재 숭실대학교 IT대학 컴퓨터학부 부교수
관심분야: 컴퓨터 비전, 형태인식, 생체인식, 증강현실, 신호처리 등