

# 현실감 있는 3차원 얼굴 애니메이션을 위한 실시간 표정 제어

## A Realtime Expression Control for Realistic 3D Facial Animation

김 정 기\*      민 경 필\*\*      전 준 철\*\*\*      최 용 길\*\*\*\*  
Kim, Jung-Gi    Min, Kyong-Pil    Chun, Jun-Chul    Choi, Yong-Gil

### 요 약

본 논문에서는 실시간으로 입력되는 동영상으로부터 영상 내에 존재하는 사람의 얼굴 및 얼굴 특징점들을 자동으로 추출한 후, 추출된 정보를 이용하여 3차원 얼굴 모델의 표정을 실시간으로 제어함으로써 현실감 있는 얼굴 애니메이션 처리가 가능한 새로운 방법을 제시한다. 입력 영상의 각 프레임으로부터 얼굴을 효과적으로 추출하기 위해 기존에 일반적으로 사용되는 색상 공간을 이용한 파라미터 검출 방법에 대변되는 새로운 비파라미터 검출 방법을 제시하였다. 기존의 파라미터 검출 방법은 일반적으로 얼굴의 피부 색상분포를 가우지언 형태로 표현하며 특히 주변조명의 변화 및 배경 영상 등에 민감하게 반응하므로 정확한 영역의 검출을 위한 추가적 작업을 필요로 한다. 이러한 문제점을 효과적으로 해결하기 위하여 본 논문에서는 Hue와 Tint 색상 성분에 기반을 둔 새로운 스킨 색상 공간을 제시하고 모델의 분포특성을 직선 형식으로 표현하여 얼굴검출 시 발생하는 오류를 축소시킬 수 있었다. 또한, 검출된 얼굴 영역으로부터 정확한 얼굴특성 정보를 추출하기 위하여 각 특징영역에 대한 에지검색 결과와 얼굴의 비율 비를 적용하여 효과적으로 얼굴의 특징 영역을 검출하였다. 추출된 얼굴 특징점 변화 정보는 3차원 얼굴 모델의 실시간 표정 변화에 적용되며, 보다 실감 있는 얼굴 표정을 생성하기위하여 사람의 근육 정보와 근육의 움직이는 방법을 나타내는 Waters의 선형 근육 모델에 새로운 근육 정보들을 새롭게 추가함으로써 확장 적용하였다. 실험결과 제안된 방법을 이용하여 실시간으로 입력되는 대상의 얼굴표정을 3차원 얼굴 모델에 자연스럽게 표현할 수 있다.

### Abstract

This work presents a novel method which extracts facial region and features from motion picture automatically and controls the 3D facial expression in real time. To extract facial region and facial feature points from each color frame of motion pictures a new nonparametric skin color model is proposed rather than using parametric skin color model. Conventionally used parametric skin color models, which presents facial distribution as gaussian-type, have lack of robustness for varying lighting conditions. Thus it needs additional work to extract exact facial region from face images. To resolve the limitation of current skin color model, we exploit the Hue-Tint chrominance components and represent the skin chrominance distribution as a linear function, which can reduce error for detecting facial region. Moreover, the minimal facial feature positions detected by the proposed skin model are adjusted by using edge information of the detected facial region along with the proportions of the face. To produce the realistic facial expression, we adopt Waters's linear muscle model and apply the extended version of Water's muscles to variation of the facial features of the 3D face. The experiments show that the proposed approach efficiently detects facial feature points and naturally controls the facial expression of the 3D face model.

☞ Keyword : keyword: Facial Expression, Nonparametric Skin Model, Facial Animation, Facial Feature Detection

## 1. 서 론

\* 준 회 원 : 링크웨어 연구원

reds20@nownuri.net

\*\* 정 회 원 : 경기대학교 전자계산학과 강사

buffalo21@hotmail.com

\*\*\* 종신회원 : 경기대학교 전자계산학과 교수

jcchun@kyonggi.ac.kr

\*\*\*\* 정 회 원 : 경기대학교 경영학부 교수

yychoi@elim.net

[2005/10/07 투고 - 2005/10/25 심사 - 2006/2/17 심사완료]

컴퓨터를 이용한 사람의 얼굴에 대한 연구는 일반 영상이나 동영상 분석을 통해 얼굴 또는 얼굴의 특징 영역을 검출하고 표정을 인식하는 연구를 시작으로 최근에는 3차원 아바타 생성, 원격 회의 그리고 인간과 컴퓨터 간의 상호작용

(Human and Computer Interaction) 등과 같은 다양한 분야로 적용 범위가 확대되고 있다. 특히 영상에서 획득된 얼굴 및 특징 영역 정보를 이용하여 3차원 얼굴 모델의 표정을 제어하는 방법에 대한 많은 연구들이 진행되고 있다 [1,2,4,5].

실시간 동영상 분석을 통해 획득한 2차원 정보를 이용하여 3차원 얼굴 모델을 제어하는 방법은 크게 두 가지로 구분될 수 있다. 첫 번째는 광학식 모션 캡처 방법으로 입력 대상이 되는 얼굴에서 표정의 변화를 감지할 수 있는 곳(눈, 입 등)에 마커를 부착한 후, 동영상에서 마커의 움직임을 추적하는 방법이다[2,5]. 마커는 입력 대상의 표정에 따라 변화하게 되고, 이 정보를 3차원 얼굴 모델에 적용함으로써 대상과 아주 유사한 표정을 나타낼 수 있다는 장점이 있다. 그러나 광학식 모션 캡처 방법은 사람의 얼굴에 마커를 직접 부착해야 하는 불편성과 마커의 움직임을 추적하는 추가 장비가 필요하다는 단점이 존재한다. 두 번째로는 동영상으로부터 직접 얼굴 정보를 분석하여 얼굴 및 얼굴의 특징 영역들을 추출하여 3차원 모델에 적용하는 방법이다. 이러한 방법은 입력 동영상으로부터 얼굴과 얼굴 특징 영역을 추출하는 단계와 얼굴 모델을 제어해서 애니메이션 처리를 하는 단계로 나눌 수 있다.

얼굴 및 얼굴 특징 영역을 추출하는 방법은 영상기반 방법(image-based method)과 특징기반 방법(feature-based method)으로 분류된다. 영상기반 접근방법은 영상의 특정 부분이 아닌 영상 전체를 이용하는 방법으로, 신경망, 통계적 방법 그리고 주성분 분석과 같은 선형 부분 공간방법을 이용해서 얼굴과 특징 영역을 추출하는 방법을 의미한다[6,7,8]. 특징기반 방법은 얼굴의 변하지 않는 특징들을 이용하는 방법으로 색상, 에지, 형태, 밝기값 등을 이용하여 얼굴과 특징 영역을 추출하는 방법이다[1,9,10,11]. 특히 얼굴의 색상 정보를 이용한 특징기반 방법은 연산 속도가 빠르고, 비교적 정확한 얼굴영역을 추출할 수 있다

는 특성이 있다. 색상을 이용한 추출 방법은 색상 공간으로부터 피부영역과 피부영역이 아닌 곳으로 분리함으로써 얼굴을 추출한다. 얼굴 검출을 위해 이용되는 기존 색상 공간으로는 CbCr 색상 공간[1,9], HS 색상 공간[13], 그리고 TS 색상 공간[10] 등이 우수한 검출 성능을 보장한다.

얼굴 모델을 제어하여 애니메이션 효과를 나타내기 위해 사용되는 방법은 일반적으로 키프레임 애니메이션 기법, 매개변수 모델 기법 그리고 근육기반 모델링 기법 등이 이용된다. 키프레임 애니메이션 기법은 임의의 표정을 만들기 위해 중요한 키프레임에 대하여 얼굴 모델의 상태를 지정하고 이들 사이를 보간 함으로써 표정을 생성하는 방법이다[4]. 키프레임 기법은 얼굴 표정이 변화하는 처음 부분과 끝 부분의 정의를 수작업에 의존하게 되어 제작 시간이 길고 사실적인 표정 애니메이션을 생성하기가 어렵다는 단점을 지니고 있다. 매개변수 모델 기법에서는 얼굴 표정의 변화를 수십 개의 매개변수를 이용해서 정의한다[15]. 이러한 방법은 매개변수들의 값들과 실제 얼굴 표정간의 관계를 정확하게 파악하기 힘든 것이 가장 큰 단점이다. 또 다른 접근 방법인 근육기반 모델링은 사람의 표정 변화 시 움직이는 안면 근육들의 정보와 근육들의 움직이는 방법을 나타낸 근육 모델을 이용하여 실제 사람의 얼굴에 표정을 나타낼 때와 유사한 원리로 3차원 얼굴 모델의 표정을 제어하는 방법이다[16].

본 논문에서는 실시간으로 입력되는 동영상으로부터 영상 내에 존재하는 사람의 얼굴과 얼굴의 특징 영역들을 자동으로 추출한 후, 추출된 정보를 이용하여 3차원 얼굴 모델의 표정을 실시간으로 제어하는 방법을 제시한다. 일반적으로 3차원 얼굴모델의 제어는 제작시간이 오래 걸리는 수작업에 의존하거나 얼굴의 표정을 캡처 하는 장비를 사용하게 된다. 이러한 단점을 보완하기 위해서 동영상을 실시간으로 처리하여 사람의 간섭 없이 효과적으로 얼굴의 특성을 추출하는 방법으로 상대적으로 연산이 적은 특징기반 접근방

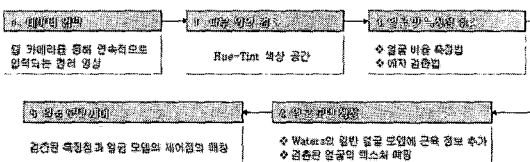
법을 사용하였다. 또한 현실감 있는 얼굴 애니메이션을 위해 Waters의 일반 얼굴 모델의 선형 근육 모델(linear muscle model)[12]을 확장 적용하였다. 기본적으로 제공되는 Waters의 3차원 얼굴 근육 모델에는 입 주변에 대한 근육 정보가 부족하여, 입을 벌리거나 웃는 등의 움직임이 부자연스럽다. 이러한 문제를 해결하기 위하여 입 주위의 근육정보인 소근(risorius), 구각하재근(triangularis)을 각각 입을 중심으로 좌우에 1개씩, 그리고 입술 중심에 이근(mentalis) 1개 등 총 5개의 근육을 새롭게 추가하여 보다 현실감 있는 애니메이션이 가능한 얼굴 모델을 생성한다.

본 논문의 구성은 다음과 같다. 2절에서는 실시간 얼굴 제어를 위해 새롭게 제안한 시스템 구조를 기술한다. 실시간 제어를 위해 얼굴 및 얼굴의 특징점을 추출할 수 있는 방법과 3차원 얼굴 모델을 제어하는 방법에 대해 기술한다. 3절에서는 제안 방법에 의한 실험 결과와 실시간 얼굴 제어 결과를 소개하며, 6절에서 본 논문에 대한 결론을 맺도록 한다.

## 2. 실시간 얼굴표정 제어 시스템

### 2.1 제안 시스템 구조

본 논문에서 제안하는 3차원 얼굴 모델의 실시간 표정 제어를 통한 애니메이션 처리 시스템의 전체 구조는 그림 1과 같다. 제안된 방법은 크게 얼굴 및 특징점 추출 과정과 얼굴 애니메이션 과정으로 나눌 수 있다.



〈그림 1〉 제안 시스템 구조

얼굴 및 특징점을 추출하기 위해서 사용되는

입력 영상은 비디오 혹은 웹 카메라를 통해 실시간으로 입력되는 컬러 동영상을 이용한다. 본 논문에서는 입력 데이터로 황인종의 피부 색상을 갖는 정면 얼굴 영상만을 입력 받는 것으로 가정한다. 얼굴 및 특징점을 추출하기 위해서 비교적 정확한 검출 결과와 빠른 처리 능력을 보장하는 특징기반 검출 방법 중 색상 정보를 이용하는 방법을 적용한다. 특히, 외부 조명의 변화에 강건하고 다양한 피부 색상 영역을 검출할 수 있도록 Hue와 Tint 성분을 기반으로 한 비파라미터 색상 공간인 HT를 적용하도록 한다. HT 색상 공간 내의 피부 영역을 모델링하기 위해서 직선 함수를 적용하여 다양한 피부색의 분포 형태를 정의한다. 제시된 색상 모델과 입력 영상의 유사도를 측정하여 입력 영상으로부터 피부 영역을 검출하며, 후처리로 불필요한 잡음을 제거하기 위하여 미디언 필터를 적용하였다. 또한 검출된 피부 영역에서 얼굴에 대한 비율 정보와 히스토그램 정보를 이용하여 눈과 입에 대한 특징 영역의 위치를 추출 하며, 보다 정확한 특징 영역의 위치를 찾기 위해 에지 연산과 이진화 연산과 같은 추가적인 연산을 적용하였다.

얼굴 애니메이션 과정에서는 근육 정보를 포함하고 있는 Waters의 일반 얼굴 모델을 이용한다. 이 과정에 이용되는 일반 얼굴 모델에는 실시간 얼굴의 표정 제어를 위해서 이전 과정에서 추출된 특징점과 정합할 수 있도록 같은 수의 특징점들을 미리 정의해놓는다. 부드러운 얼굴의 움직임을 위해서 선형 근육 모델을 이용하였으며, 기존의 Waters의 일반 얼굴 모델이 입 부분의 애니메이션이 부자연스럽다는 단점을 해결하기 위해서 턱 영역에 추가적인 근육 정보를 새롭게 삽입함으로써 실시간으로 입력되는 영상 내의 사람의 움직임과 유사한 표정을 나타내도록 하였다.

### 2.2 얼굴 피부 영역 검출

색상을 이용하여 영상으로부터 얼굴 영역을

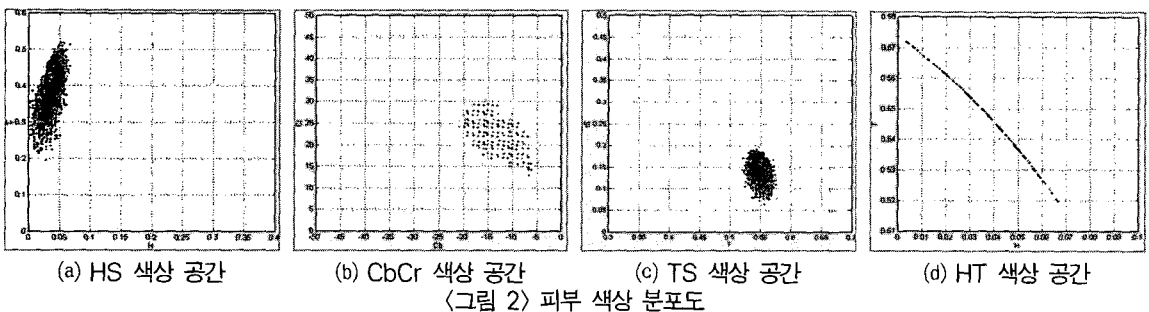
검출하기 위해서는 얼굴의 피부 색상과 영상 내 유사한 색상을 쉽게 구분할 수 있는 색상 공간을 이용한다[5]. 색상 공간을 이용하기 위해서는 피부 색상을 가장 잘 표현할 수 있는 색상 공간을 우선 선택해야하며, 피부색의 분포도, 피부 영역과 피부 영역이 아닌 영역의 중복도, 그리고 분포 영역의 모델링 방법을 고려해야한다. 적합한 색상 공간의 세 가지 선택 기준은 다음과 같다. 첫 번째, 피부색에 대한 분포는 좁은 영역에 밀집되는 형태의 분포가 적합하며, 두 번째, 피부와 피부가 아닌 색상의 중복성은 피부 영역의 검출 과정에서 오검출율을 줄일 수 있도록 최대한 중복도가 낮은 색상 공간을 선택한다. 마지막으로 피부 영역의 모델링은 분포 형태를 수식으로 표현하고 입력 데이터에 대해 피부 영역과 피부가 아닌 영역을 판단하는 방법을 간단하게 표현할 수 있는 방법이 적합하다.

실험을 위하여 얼굴 피부 색상에 대한 표본 데이터를 이용하여 색상 공간에서의 분포도를 조사한다. 사람의 얼굴 자체를 표본으로 사용했을 경우 피부색과 관련이 없는 눈이나 눈썹, 머리카락에 의해 정확한 표본의 역할을 하지 못하는 문제점을 가지게 된다. 따라서 표본 데이터 생성을 위해서 시각적 인지 능력을 충분히 가진 사람이 얼굴의 피부 영역만을 직관적으로 추출하였다. 본 논문에서는 약 600개의 얼굴 영상을 이용하며, 얼굴 영상 중에서 다양한 조명과 배경에서 촬영된 118개의 얼굴 영상을 통해 표본 데이터를 생성하였다.

얼굴 영역 검출에 이용되는 색상 공간은 CbCr, HS, 그리고 TS 색상 공간이 가장 많이 이용된다. 이러한 각 색상 공간들을 구성하는 성분들을 Cb, Cr, H, S, T, tS(HSI 색상 모델에서의 S와 TSL의 S성분을 구별하기 위해 tS로 표기함)로 분리하고, 이들을 재조합함으로써 15가지 색상 공간을 생성할 수 있다[17]. 본 논문에서는 새롭게 생성된 색상 공간 중 HT 색상 공간을 이용한다. 그림 2는 기존 색상 공간과 HT 색상 공간에서의 피부색에 대한 분포를 나타낸 것이다.

그림 2의 각 색상 공간에 분포되어 있는 피부 영역은 (a)에서는  $([0.0, 0.07], [0.2, 0.5])$  사이에 분포하며, (b)에서는  $([-20, -5], [10, 30])$ , (c)에서는  $([0.5, 0.6], [0.05, 0.2])$ , 그리고 (d)에서는  $([0.0, 0.07], [0.52, 0.57])$  사이에 분포한다. 그러므로 HT 색상 공간에서 가장 높은 밀집도를 보임을 확인할 수 있다.

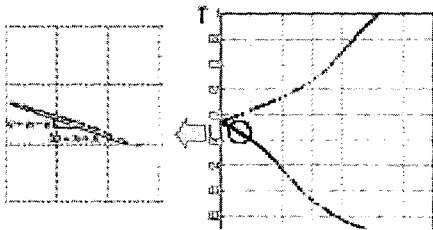
색상 공간을 이용한 피부 영역 검출의 효율성 평가를 위하여, 각 색상 공간 내에서 피부 영역과 피부가 아닌 영역간의 중복 정도를 측정하였다. 정규화 된 피부 영역과 피부가 아닌 영역의 히스토그램 교차율을 측정한 결과는 표 1과 같으며, HT 색상 공간의 경우 피부영역과 피부가 아닌 영역의 교차율이 다른 모델에 비하여 상대적으로 적게 나오는 것을 확인할 수 있다. 이처럼 히스토그램 교차율이 적다는 것은 서로 다른 두 영역을 잘 분리할 수 있음을 의미한다.



〈표 1〉 색상 공간에 따른 피부영역과 비피부영역의 교차율 비교

색상 공간	H-S	Cb-Cr	T+S	HT
중복율	0.1218	0.1311	0.1182	0.1182

피부 영역에 대한 모델링을 위해 HT 색상 공간에서의 피부 색상의 분포 형태를 분석하도록 한다. 일반 영상에 대한 색상 분포를 HT 색상 공간에 그려보면 그림 3의 (b)와 같이 직선 형태를 이루고 있다. 그러므로 전체 분포도에서 피부에 해당되는 분포(그림 3 (a))만을 분리하면 직선의 방정식을 이용하여 간단하게 수식화 할 수 있다.



(a) 피부 영역 분포도 (b) 일반 영상 분포도  
〈그림 3〉 피부색 분포의 선형 모델화

직선의 방정식은 임의의 두 점을 알고 있으면 구할 수 있으며, 그림 3의 (a)에서 볼 수 있듯이, 피부 영역에 대한 H 성분과 T 성분의 최대값과 최소값의 조합을 통해 두 개의 임의의 점을 정의할 수 있다. 식 (1)은 H와 T의 최대값과 최소값에 해당되는 두 개의 점을 지나는 직선의 방정식을 나타낸다.

$$t - T_{\max} = \frac{T_{\min} - T_{\max}}{H_{\max} - H_{\min}} (h - H_{\min}), \quad H_{\min} \leq h \leq H_{\max} \quad (1)$$

식 (1)에서 변수  $h$ 와  $t$ 는 입력영상의 특정

화소  $(x, y)$ 의 H와 T의 값을 의미하며,  $T_{\min}$ ,  $H_{\min}$ 는 T와 H의 최소값,  $T_{\max}$ ,  $H_{\max}$ 는 T와 H의 최대값을 의미한다. 얼굴영역은 입력 영상의 각 화소를 HT 색상 공간으로 변환한 후, 정의된 피부모델과 입력영상 사이의 유사도 측정(similarity measure)에 의해 검출할 수 있다. 유사도 측정을 위해서는 점과 직선 사이의 거리를 측정하는 방법을 적용하도록 한다. 임의의 점  $P(h_1, t_1)$ 과 직선  $ah + bt + c = 0$  사이의 거리는  $d = \frac{|ah_1 + bt_1 + c|}{\sqrt{a^2 + b^2}}$ 가 되므로, 식 (2)와 같이 유도된다.

각 표본 데이터에서 피부 영역의 화소값에 대한 H, T값  $(h, t)$ 와 식 (3)과의 거리를 비교해 본 결과 최대 오차는 0.003이므로, 이 값을 피부 영역 검출에 이용되는 임계치로 이용한다. 이러한 정의는 기존의 많은 색상 공간이 해당 공간의 평균값이나 분산값에 의존하여 계산량이 증가하는 문제를 해결할 수 있는 장점도 보일 수 있다.

HT 색상 공간은 앞서 설명한 적합한 색상 공간의 세 가지 선택 기준을 만족할 뿐만 아니라, 외부 조명의 변화에도 H값의 수치적 변화가 생기지 않는다는 특성을 지니므로, 조명에 대한 영향을 덜 받는 장점도 포함하게 된다. 그림 4는 디지털 카메라에 의해 촬영된 영상과 웹 카메라에 의해 촬영된 동영상으로부터 얼굴 피부 영역을 검출한 결과이다.

정확한 얼굴 추출을 위해 발생한 잡음들을 제거해야 한다. 잡음은 조명, 유사색, 그리고 캡처 장비에서 발생하는 신호적 오류로 인해 발생하게 된다. 이러한 잡음들은 추출된 얼굴 영역에 비해 화소의 크기가 작으며, 응집되어 있지 않고 무작위로 발생하게 된다. 이러한 특성을 고려하

$$d(h, t) = \frac{(T_{\max} - T_{\min})h + (H_{\max} - H_{\min})t + (T_{\min} - T_{\max})H_{\min} + (H_{\min} - H_{\max})T_{\max}}{\sqrt{(T_{\max} - T_{\min})^2 + (H_{\max} - H_{\min})^2}} \quad (2)$$

여, 비선형 필터인미디어 필터를 사용하여 잡음을 제거한다.

다양한 영상에 대해 기존 색상 공간인 CbCr, TS, HS 색상 공간과 HT 색상 공간을 이용한 피부 영역 검출 결과는 그림 4와 같다.

피부 영역의 검출 결과의 정확성을 평가하기 위해서 기존 색상 공간 중 가장 성능을 보이는 것으로 증명된 TS 색상 공간[5]과 HT 색상 공간에 의한 검출 결과를 비교하였다. 본 연구에 사용되는 실험 데이터는 조명과 주변 환경이 다른 상황에서 획득한 약 600개의 얼굴 영상을 6개의 얼굴 집단으로 분류해서 이용한다. 1집단은 일반 영상을 포함하고 있으며, 2집단은 어두운 조명을 주고 촬영한 영상 집단, 3집단은 사진관에서 자연광과 유사한 조명을 주고 촬영한 영상 집단, 4집단은 피부색과 유사한 배경을 포함한 영상 집단, 5집단은 화장이 짙은 여성들을 촬영한 영상 집단 그리고 6집단은 얼굴에 과도한 조명을 준 영상 집단으로 구분된다. 표 2는 피부 영역의 검출율을 정의한 것으로서, 모든 집단에 대해 기존의 TS 색상 공간보다 높은 검출율을 보여주며, 모든 과도한 조명이 들어간 6집단을 제외하고는 높은 검출 성능을 보장한다.

〈표 2〉 TS 과 HT 색상 공간에 의한 얼굴 검출률

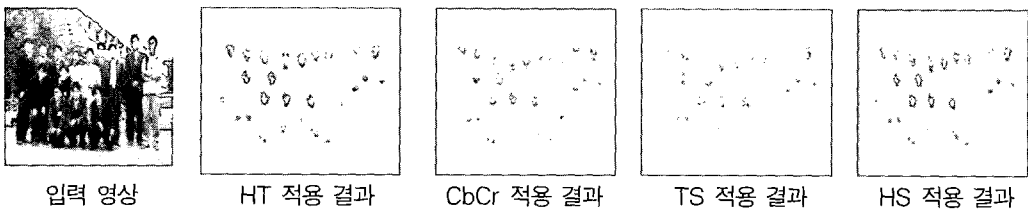
훈련 집단	TS 색상 공간	HT 색상 공간
1 집단	60.0 %	72.5 %
2 집단	65.0 %	90.0 %
3 집단	73.3 %	83.3 %
4 집단	66.6 %	93.0 %
5 집단	70.0 %	100 %
6 집단	23.0 %	42.0%

### 2.3 얼굴 특징점 추출

검출된 피부 영역으로부터 얼굴 영역만을 선택하기 위해서 얼굴의 특징점을 분석한다. 검출된 피부 영역 내에 눈, 눈썹, 입과 같은 특징 영역이 존재할 경우 얼굴로 간주할 수 있으므로, 검출된 얼굴 후보 영역에서 특징점을 검출한다. 검출된 피부 영역을 얼굴로 간주할 경우 검출된 영역에서는 다음과 같은 특징을 정의할 수 있다. 첫째, 일반적으로 얼굴에서 눈, 눈썹, 입 영역은 피부색과 다른 색으로 구성되어 있기 때문에 피부색 추출 결과에서 제외 되어 영역 내에 구멍이 생성된다. 두 번째, 얼굴의 기본적인 비율 정보에 따라 얼굴의 두 눈은 얼굴 상단에 위



(a) 일반 영상에 대한 검출 결과



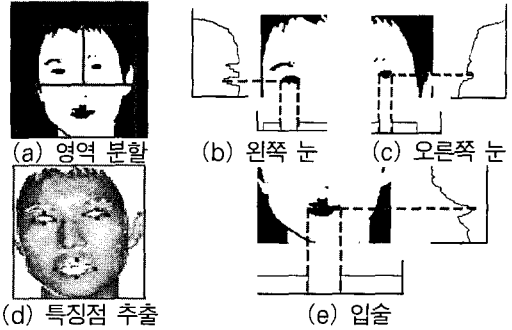
(b) 복잡한 배경으로부터 피부 영역 검출 결과

〈그림 4〉 다양한 색상 공간을 이용한 피부 영역

치하고, 입은 하단에 위치한다. 즉, 얼굴을 상좌·상우·하단과 같이 세 부분으로 나누어서 얼굴의 특징 영역들이 추출한 후 특징 영역의 후보 영역들을 선별한다. 위 부분에 나누어진 두 개의 영역에서는 눈과 눈썹 후보 영역이 검출되며, 아랫부분으로 나누어진 영역에서는 입 후보 영역이 검출되게 된다.

구분된 각 영역에 대하여 각 열들의 합에 대한 누적 그래프를 생성하면 그래프에서 발생하는 최소값을 검출할 수 있으며, 검출된 최소값에 대한 행의 값을 표현하는 그래프를 작성하여 그래프에서의 최소값을 찾아낼 수 있다. 이러한 최소값들의 위치 정보를 분석하면, 그림 5와 같이 얼굴 영역에서 눈과 입 영역에 대한 위치를 검출할 수 있다.

얼굴의 특징점들의 보다 정확한 검출과 색상 비교에 의해 검출되지 않는 턱과 같은 특징점들을 검출하기 위해서 이진화 연산과 에지 검출 연산을 적용한다. 사람의 얼굴 구조는 눈과 입의 위치 정보를 기준으로 특정 비율을 갖고 있으므로, 검출된 눈과 입의 위치를 기준으로 다른 특징 후보 영역이 존재할 수 있는 후보 영역들을 블록으로 설정할 수 있다. 먼저 특징 영역들 중 눈과 입술과 같이 피부색과의 구별이 명확한 경



〈그림 5〉 얼굴의 특징점 검출 과정

우, 명암도 차이를 이용한 분류 연산인 이진화 연산을 적용하여 눈과 입술을 얼굴에서 분리할 수 있다. 그러나 검출된 눈과 입술 영역은 잡음이나 조명에 의해 오차가 생길 수도 있다. 이러한 문제는 각 특징 영역과 특징 후보 영역에 대한 블록의 에지 정보를 이용하여 해결할 수 있다. 에지 연산으로 많이 사용되고 있는 3×3 소벨 연산자를 각 블록에 적용하고, 에지에 대한 누적 그래프를 그려줌으로써 보다 정확한 특징점의 위치와 턱 끝과 아래턱에 해당하는 추가적인 특징점들을 검출할 수 있다. 본 연구에서는 총 20개의 특징점을 검출한다. 그림 6은 제안 방법에 의하여 얼굴의 특징 영역들이 검출되는 과정을 차례로 도식화한 그림이다.



(a) 디지털 카메라에 의해 촬영된 영상  
(b) 웹 카메라에 의해 촬영된 영상  
(왼쪽부터 : 입력 영상, 이진화 영상, 특징점 후보 검출 결과, 특징점 후보 영역 설정, 특징점 검출 결과)

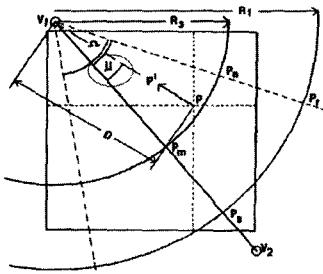
〈그림 6〉 얼굴의 특징점 검출 과정

그림 7은 다양한 얼굴 영상에 대한 특징점 검출 결과이다.



### 2.4 3차원 얼굴 모델의 표정 제어

본 논문에서는 사람의 안면 근육의 움직임을 모델화한 선형 근육 모델 [7]을 이용하여 실시간 동영상의 분석을 통해 추출한 얼굴의 특징 영역들의 움직임을 자연스럽게 3차원 얼굴 모델에 나타나게 한다. 추출된 특징점들을 나타낼 3차원 얼굴 모델은 Waters의 일반 얼굴 모델을 사용하도록 한다. Waters의 얼굴 모델은 기본적으로 512개의 점과 18개의 안면 근육으로 구성되며, 3차원 얼굴 모델의 제어는 실제 사람의 얼굴의 근육 움직임을 반영하는 선형 근육 모델을 이용하였다. 그림 8은 선형 근육 모델의 움직임과 이러한 움직임이 주변의 피부에 어떠한 영향을 미치는지를 나타내고 있다.



〈그림 8〉 선형 근육 모델

$v_1$ 과  $v_2$ 는 근육의 시작과 끝을 나타내며 그들을 연결한 직선이 선형 근육이다. 임의의 노드  $p$  (피부 : vertex) 가  $p'$ 로 이동할 때  $v_1, p, p_s$ 범위 내에서 벡터  $p$ 와  $v_1$ 을 따라서  $v_1$ 방향으로 진행한다, 식 (3)과 같이 표현된다.

$$p' = p + akr \frac{pv_1}{|pv_1|}, a = \cos(\mu) \quad (3)$$

$a = \cos(\mu)$ 는 벡터( $v_1, v_2$ )와  $(v_1, p)$ 사이각을 나타내고,  $r$ 은  $p$ 가 포함된 영역에 따라 식 (4)와 같은 값을 가지게 된다.

$$r = \begin{cases} \cos\left(\frac{1-D}{R_s}\right) & \text{for } P \text{ inside } V_1P_nP_m \\ \cos\left(\frac{D-R_s}{R_f-R_s}\right) & \text{for } P \text{ inside } P_nP_mP_sP_R \end{cases} \quad (4)$$

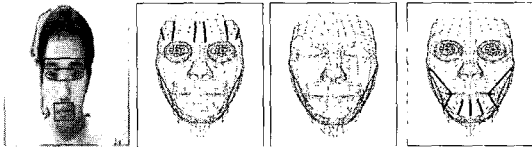
식 (4)를 선형 근육 모델에 직접 적용하여 근육의 움직임을 제어할 수 있다.

추출된 특징점들을 나타낼 3차원 얼굴 모델인 Waters 모델은 앞서 지적한바와 같이 입술 주변의 근육 정보가 부족하여 입 주변의 애니메이션에서 특히 자연스럽게 못하다. 이런 문제를 해결하기 위해 본 논문에서는 의학적 근육 정보를 참고하여 각각 좌우에 소근과 구각 하계근을 1개씩, 턱 중앙에 이근 1개 등 총 5개의 근육을 추가하여 입 주변의 어색한 애니메이션 효과를 보완하였다.

입력 대상의 표정을 3차원 얼굴 모델에 나타내기 위해서는 대상의 특징 영역들의 움직임을 3차원 얼굴 모델에 반영해야 한다. 이때, 각 특징 영역들의 움직임을 반영하는 근육들은 그림 9와 같다.

눈썹의 움직임은 6개의 이마 근육을 사용하여 눈썹을 올린 상태를 나타낼 수 있고, 눈 영역은 11개의 제어점을 회전 이동 하여 눈을 감거나 뜬





(a)특징영역 (b)이마근육 (b)눈제어점 (c)입 근육  
 <그림 9> 특징점에 대응되는 근육과 제어점

애니메이션을 나타낸다. 마지막으로 입 영역은 기존의 4개의 근육정보에 5개의 추가 근육정보를 더하여 총 9개의 근육정보를 사용하여 입을 벌리거나 웃는 애니메이션 효과를 나타낸다. 그림 10과 11은 얼굴의 특징 영역들의 변화에 따라 나타나는 3차원 얼굴 모델의 표정 변화를 앞서 정의한 근육 정보를 이용해 생성한 결과이다.

### 3. 실험 결과

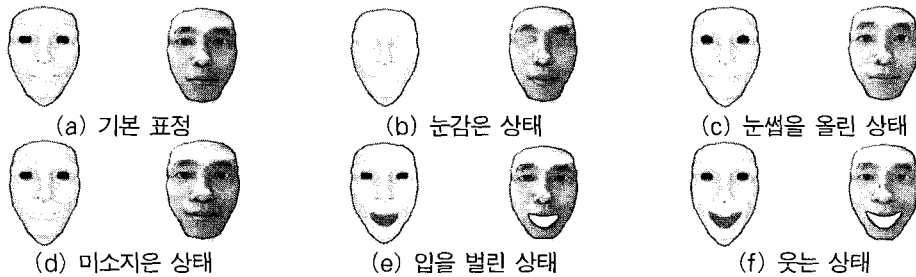
제안된 표정 제어 시스템은 Visual C++ 6.0 기반 하에서 개발되었고, 윈도우에서 제공하는

모듈을 사용하기 위해 MFC 라이브러리를 사용한다. 웹캠으로부터 영상을 입력받기 위해 VFW (Video For Window) 라이브러리를 사용하고, 3차원 모델을 표현하기 위해 OpenGL 라이브러리를 사용한다.

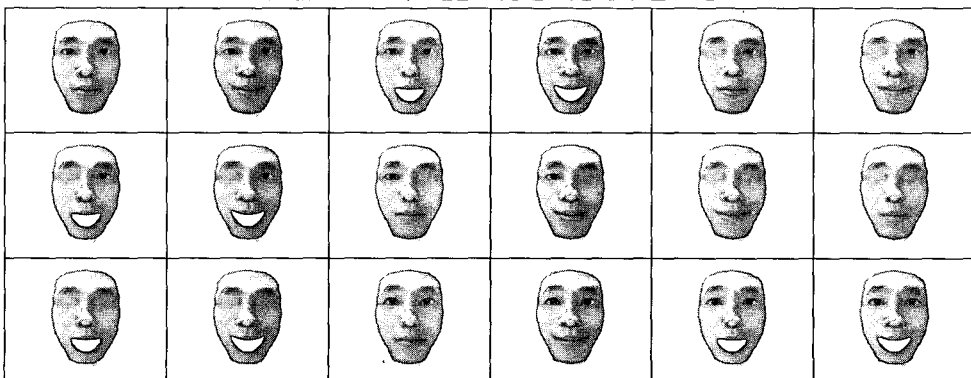
표 3은 추출된 특징점들의 변화 정도에 따른 모델의 표정 형태를 정의한 것이다. 얼굴의 변화 정도를 살펴보기 위해서 무표정한 상태의 얼굴로부터 추출된 특징 영역들의 값을 초기값 (Init)으로 설정하고, 매 1초 간격으로 특징 영역들의 크기 정보와 위치 정보를 비교하여 어떠한 표정을 짓고 있는지를 정의하였다.

표 3에서 정의된 얼굴의 특징 영역들의 크기 변화와 위치 변화를 3차원 모델에 적용하기 위해서 사용되는 특징 영역들의 정보는 다음과 같다.

- LEye.Height : 1초(24Frame)동안의 평균적인 왼쪽 눈이 높이



<그림 10> 근육모델을 적용한 다양한 얼굴 표정



<그림 11> 다양한 표정 제어 결과

〈표 3〉 특징 영역 정보와 모델 표정간의 관계

모델의 표정 상태	특징 영역들의 변화
눈을 감은 상태	$LR.Eye.H < Init/2$
눈을 뜬 상태	$LR.Eye.H \geq Init$
입을 벌린 상태	$Lip.H > Init+(Init*0.4)$
웃는 상태	$Lip.H > Init+(Init*0.2)$ $Lip.W > Init+(Init*0.5)$
입을 닫은 상태	$Lip.HW \leq Init$
눈썹	$LR.Eye.H \geq Init+(Init*0.2)$ $Brow.Y - Init \geq 15$

- REye.Height : 1초(24Frame)동안의 평균적인 오른쪽 눈이 높이
- Lip.Width : 1초(24Frame)동안의 평균적인 입술의 너비
- Lip.Height : 1초(24Frame)동안의 평균적인 입술의 높이
- Brow.XY : 1초(24Frame)동안의 눈썹의 평균적인 위치 정보

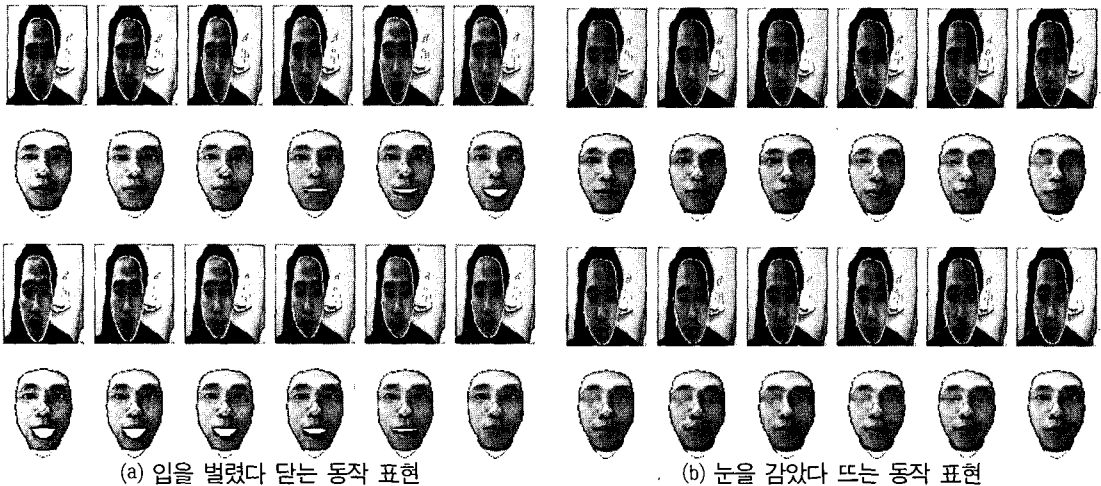
이상의 특성 정보를 이용하여 실시간으로 추적된 입력영상 내 사람의 표정을 3차원 얼굴모델에 반영한 3차원 얼굴모델의 애니메이션 결과는 그림 12와 같다. 그림 12는 8 Frame 간격으

로 캡처된 결과이며, 1초 동안의 평균적인 움직임에 관한 정보를 바탕으로 애니메이션 되고 있다. 그러므로 입력 동영상에 대한 표정 변화에 대한 얼굴 모델의 대응 반응은 1초 후에 이루어지며, 3장(24 frame)이 지난 후에 표정에 변화가 이루어짐을 확인할 수 있다.

#### 4. 결론

본 논문에서는 모션 캡처와 같은 장비를 이용하지 않고, 얼굴 표정을 제어하여 현실감 있는 얼굴 애니메이션이 가능한 방법을 제안하였다. 특히, 입력 동영상의 분석을 통해 영상 내에 존재하는 사람의 표정을 실시간으로, 사용자의 간섭 없이 3차원 얼굴 모델에 반영하여 얼굴모델의 표정을 제어하는 방법을 제시하고 실험을 통해 그 방법의 효율성을 입증하였다.

실시간 처리를 위해 얼굴 및 얼굴의 특징점을 검출하기 위해서 주변의 조명변화에 강건한 비파라미터 HT 색상 공간을 적용함으로써 기존의 색상 모델에 비하여 효과적으로 빠른 시간에 얼굴영역을 검출할 수 있었다. 보다 정확한 특징점 검출을 위해 얼굴의 비율 측정법과 에지 검출



〈그림 12〉 3차원 얼굴모델의 표정제어 결과

기법을 추가로 적용하였다. 3차원 얼굴모델의 표정제어 시, 얼굴 근육의 움직임의 사실성을 높이기 위하여 Waters의 선형 근육 모델에 추가 근육 정보를 확장 적용하여 실험 결과에서 제시된바와 같이 부드러운 애니메이션 효과를 얻을 수 있었다.

향후 연구 과제로는 보다 많은 얼굴의 특징영역에 대한 자동 추출을 통해 근육 정보의 활용을 확대시키는 방법을 모색해야 할 것이다. 현재 사용된 기본 얼굴모델은 입력 대상자와는 골격이 다른 형태이기에 입력대상자와의 유사성이 부족하다. 이러한 문제를 해결할 수 있도록 3차원 얼굴 모델의 변형을 통해 기본 얼굴 모델이 입력영상의 얼굴과의 유사성을 향상시키는 개선이 필요하다. 또한 제안 방법은 얼굴의 정면 영상에 대한 검출 기법에 적합하므로, 다양한 각도에서 촬영된 영상으로부터의 검출이 가능한 방법으로 확대해야 할 것이다.

### 참 고 문 헌

- [1] E. Hjelmas and B. K. Low. "Face detection: A survey," Computer Vision and Image Understanding, 2001.
- [2] H. Pyun, Y. Kim, W. Chae, H. Y. Kang and S. Y. Shin, "An Example-based Approach for facial expression cloning." ACM Symposium on Computer Animation, 2003.
- [3] B. Fleming, D. Dobbs, "Animating facial features and expressions" Jenifer L. Niles, MA:Charles River Media, 1999.
- [4] F. Pighin, J. Hecker, D. Lischinski, R. Szeliski, and D. H. Salesin, "Synthesizing realistic facial expressions from photographs." SIGGRAPH'98 Proceedings, pp. 231-242, 1998.
- [5] J. Noh and U. Neumann, "Expression cloning," Proceedings of SIGGRAPH 2001, pp 271-288, 2001.
- [6] C. Burges, "A tutorial on support vector machines for pattern recognition," Data Mining & Knowledge Discovery, vol. 2, no. 2, pp 121-167, 1998.
- [7] H. A Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Trans. Pattern Anal. Mach. Intell. Vol. 20, no 1. pp. 23-28, 1998
- [8] M. Turk and A. Pentland, "Eigenfaces for Recognition," Journal of Cognitive Neuroscience, Vol. 3, pp. 71-86, 1994.
- [9] R. L. Hsu, M. A. Mottaleb, and A. K. Jain. "Face detection in color images." IEEE Trans. Pattern Analysis and Machine Intell.,24: pp. 696-706, 2002.
- [10] J. C. Terrillon, M. N. Shirazi, H. Fukamachi, and S. Akamatsu, "Comparative performance of different skin chrominance models and chrominance spaces for the automatic detection of human faces in color images," Proc. IEEE Int'l Conf. on Face and Gesture Recognition, pp. 54-61, 2000.
- [11] D. Maio and Maltoni, "Real-time face location on gray-scale static images," Pattern Recognition, vol. 33, no. 9, pp. 1525-1539, 2000.
- [12] K. Waters, "A Muscle Model for Animation Three-Dimensional Facial Expression," Computer Graphics, Vol, 21, No. 4, July, 1987.
- [13] K. Sobottka, I. Pitas, "Segmentation and Tracking of Faces in Color Images," Proceedings of the Second International Conference on Automatic Face and

- Gesture Recognition, pp. 236-241, 1996.
- [14] F. I. Parke. "Computer generated animation of faces," Masters thesis, University of Utah, Salt Lake City, UT, UTEC-CSc-72-120, 1972.
- [15] P. Ekman, and W. Friesen, "Facial Action Coding System," Consulting Psychologists Press, 1978.
- [16] D. Terzopoulos and K. Waters, "Physically-based Facial Modeling, Analysis, and Computer Animation," pp. 73-80, 1990.
- [17] K. Min, J. Chun, G. Park "A Nonparametric Skin Color Model for Face Detection from Color Images," PDCAT, pp. 115-119, 2004.

## ● 저자 소개 ●



### 김 정 기

2003년 경기대학교 전자계산학과 졸업(학사)  
2005년 경기대학교 대학원 전자계산학과 졸업(석사)  
2005년 2.~현재 톱크웨어 연구원  
관심분야 : 영상처리, 컴퓨터 비전, GIS.  
E-mail : reds20@nownuri.net



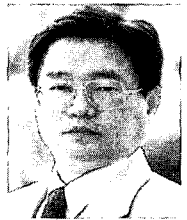
### 민 경 필

1996년 경기대학교 수학과 졸업(학사)  
1998년 경기대학교 대학원 전자계산학과 졸업(석사)  
2005년 경기대학교 대학원 전자계산학과 졸업(박사)  
2005년 8.~현재 경기대학교 전자계산학과 강사  
관심분야 : 3D Face Modeling, Face Recognition  
E-mail : buffalo21@hotmail.com



### 진 준 철

1984년 중앙대학교 전자계산학과 졸업(학사)  
1986년 중앙대학교 대학원 전자계산학과 졸업(석사)  
1992년 The Univ. of Connecticut 컴퓨터공학과 졸업(석사)  
1995년 The Univ. of Connecticut 컴퓨터공학과 졸업(박사)  
2001년 미시건 주립대학 PRIP(Pattern Recognition. & Image Processing) 객원교수  
1995~현재 경기대학교 전자계산학과 교수  
관심분야 : 3D Face Modeling & Animation, Vision-Based HCI  
E-mail : jcchun@kyonggi.ac.kr



### 최 용 길

1976년 고려대학교 무역학과 졸업(학사)  
1985년 미시건 주립대학 국제경제학 졸업(석사)  
1988년 조지아대학교 경영교육학과 졸업(석사)  
1992년 조지아대학교 경영교육학과 졸업(박사)  
2001~2002년 Texas A&M 주립대학 국제경영학과 객원교수  
1997~현재 경기대학교 경영학부 교수  
관심분야 : 전자상거래, 컴퓨터보안  
E-mail : yychoi@elim.net