

3차원 개인 얼굴 모델 자동 생성

(Automatic Generation of the Personal 3D Face Model)

咸尙續 * , 金炯坤 *

(Sang-Jin Ham and Hyoung-Gon Kim)

요 약

본 논문에서는 2차원 칼라 동영상으로부터 3차원 개인 얼굴 모델을 자동 생성하는 효율적인 방법을 다루었다. 복잡한 배경이 포함된 영상에서 얼굴 영역을 안정적으로 추출하기 위하여 피부 색상 분포에 근거한 색상 움직임 추출 방법을 이용하였다. 검출된 얼굴 영역 내에서 색상 정보와 경계선 정보를 활용하여, MPEG-4의 SNHC(Synthetic-Natural Hybrid Coding) 에서 제안하고 있는 개개인의 얼굴 특성을 표현하는 31개의 얼굴 특징점 파라메타(Facial Description Parameter : FDP)를 자동 추출하였다. 추출된 2차원 얼굴 특징점을 1038 개의 삼각형 메쉬로 이루어진 3차원 일반 얼굴 모델(Generic model)에 적용시켜 변형함으로써 개개인의 얼굴에 해당하는 모델을 자동 생성하였다. 제안하는 알고리즘은 컬러 동영상에서 배경의 복잡성, 얼굴 크기 등에 상관없이 정면상에 가까운 경우 안정적인 특징점을 추출하였으며, 펜티엄 PC에서 약 2초 이내에 개개인의 얼굴 모습에 유사한 얼굴 모델을 생성할 수 있었다.

Abstract

This paper proposes an efficient method for the automatic generation of personalized 3D face model from color image sequence. To detect a robust facial region in a complex background, moving color detection technique based on the facial color distribution has been suggested. Color distribution and edge position information in the detected face region are used to extract the exact 31 facial feature points of the facial description parameter(FDP) proposed by MPEG-4 SNHC(Synthetic-Natural Hybrid Coding) adhoc group. Extracted feature points are then applied to the corresponding vertex points of the 3D generic face model composed of 1038 triangular mesh points. The personalized 3D face model can be generated automatically in less than 2 seconds on Pentium PC.

I. 서 론

사람 얼굴에 대한 연구는 인간 심리학 등의 분야에서 상당히 오래 전부터 연구되어 왔고, 최근에는 영상 처리 및 컴퓨터 비전, 컴퓨터 그래픽스 등의 기술을 이용하여 사람의 얼굴 및 표정 인식, 얼굴 모델 기반 부호화, 가상 생명체 등의 연구에서 활발히 진행 중이다^{[1][2]}. 특히 사람의 얼굴은 개인 식별, 성별 및 심

리 상태 등을 가장 잘 나타내는 중요한 정보를 포함하기 때문에 사람과 사람간 또는 컴퓨터와 사람간의 의사 소통에 중요한 정보를 제공한다. 최근에는 얼굴의 3차원 모델을 생성하고 이를 가상 공간에서 애니메이션 하는 대화형 동영상 통신 기술이 개발 중이며, 이러한 환경에서 통신 및 기능의 표준화를 위한 작업이 MPEG(Moving Picture Experts Group)-4의 SNHC 그룹에서 진행 중이다^[3].

3차원 모델을 생성하기 위한 기존의 방법들은 데이터의 입력 방법에 따라 3차원 입력 데이터를 이용하는 방법과 2차원 입력 데이터를 이용하는 방법으로 나눌 수 있다^[4]. 3차원 입력 데이터를 이용하는 방법은 다

* 正會員, 韓國科學技術研究院 映像미디어研究센터
(Imaging Media Research Center, KIST)
接受日字:1998年5月28日, 수정완료일:1998年9月25日

시 표면 정보만을 이용하는 방법과 내부의 정보까지 이용하는 부피 방식이 있다. 표면 3차원 정보를 이용하는 방법에는 레이저 패턴을 물체 표면에 투영하여 깊이 정보를 구하는 3D scan digitizer 방식과, 실제 모형에 mesh 패턴을 그린 후 electro-magnetic 3D digitizer로 데이터를 구하는 방법 등이 있다. 부피 3차원 정보를 이용하는 경우 CT(Computer Tomography)나 MRI(Magnetic Resonance Imaging)등의 의료 영상 장비를 이용하며, 물체 내부의 특성을 반영할 수 있어 얼굴에 이용하는 경우 근육과 골격에 관한 정보를 얻을 수 있으나 고가의 장비들이 필요하고 표면정보를 정확히 구하기 어려운 단점들이 있다. 2차원 입력 데이터를 이용하여 모델을 생성하는 방법에는 스테레오 영상을 이용하여 얼굴의 깊이 정보를 알아내는 방법과, 일반 얼굴 모델(Generic face model)을 이용하는 방법이 있다. 스테레오 영상을 이용하는 경우, 정확한 3차원 모델을 만들기 위해서는 잘 정렬된 여러 대의 카메라를 이용해야 하며 계산량이 매우 큰 단점이 있다. 일반 얼굴 모델을 이용하는 방법은 사람의 얼굴형은 보통 일정한 형태를 가지고 있다는 전제하에 2차원 영상상의 얼굴에서 특징점들을 추출하고 이를 일반 모델에 적용하여 3차원 얼굴 모델을 생성시키는 방법이다. 이러한 방법으로 생성되는 얼굴 모델은 정확도는 떨어지나 비교적 빠른 모델 생성 처리가 가능하여 화상 회의, 영상 통신 등에 쉽게 적용시킬 수 있는 특징이 있다^{[5] [6] [7]}. 복잡한 배경에 정면상의 얼굴이 포함되어 있는 2차원의 칼라 영상에서 얼굴 색상 정보를 이용하여 얼굴 영역을 분리해 내는 연구는 여러 연구에서 시도되었다^{[8] [9]}. 그러나 배경에 살색이 있는 일반적인 영상의 경우 적용에 어려움이 있다. 이러한 점을 해결하기 위한 방안으로 거리 정보를 사용하여 물체를 분리하고 분리된 물체에 대해 색상정보를 이용해서 얼굴을 검출하는 방안이 제안되었으나^[10] 스테레오 영상을 사용하므로 시스템 구성이 복잡하여지는 단점이 있었다.

본 논문에서는 일반 얼굴 모델을 이용하여 개개의 3차원 얼굴 모델을 자동적으로 생성하는 방법을 제시한다. 두 장의 연속되는 영상으로부터 움직이는 살색영역을 검출하는 방법을 제시하고, 검출된 얼굴 영역 내에서 색상 정보와 경계선 정보를 활용하여, MPEG-4의 SNHC 에서 제안하고 있는 개개의 얼굴 특성을 표현하는 43개의 얼굴 특징점 파라메타

(Facial Description Parameter : FDP) 중 2차원 영상에서 표시 가능한 31 개의 특징점을 자동 추출하였다. 추출된 2차원 얼굴 특징점들을 1038 개의 삼각형 메쉬로 이루어진 3차원 일반 얼굴 모델에 적용시켜 변형함으로써 개개의 얼굴에 해당하는 모델을 자동 생성하였다. 제안하는 알고리즘은 컬러 동영상에서 배경의 복잡성, 얼굴 크기 등에 상관없이 정면상에 가까운 경우 안정적인 특징점을 추출하였으며, 펜티엄 PC 에서 약 2초 이내에 개개의 얼굴 모습에 유사한 얼굴 모델을 생성할 수 있었다.

본 논문의 구성은 다음과 같다. 서론에 이어 II장에서는 칼라 동영상에서 얼굴 색상 움직임 정보를 이용한 얼굴 영역 및 얼굴 요소 영역 검출 방법에 대하여 설명하고, III장에서는 검출된 얼굴 요소 영역 내에서 개개의 모델 생성을 위해 필요한 특징점을 추출하는 방법 및 이를 이용해서 일반 얼굴 모델을 개인 얼굴 모델로 변화하는 방법에 대해 기술한다. 제시한 방법의 검증을 위한 여러 영상에 대한 실험 방법 및 결과에 대해 IV장에서 보이고, V장에서 결론을 맺는다.

II. 피부 색상 움직임 정보를 이용한 얼굴 요소 영역 검출

본 장에서는 일반적인 배경이 있는 칼라 영상에서 얼굴 요소에 해당하는 영역을 검출하는 방법에 대해 설명한다. 기본적으로는 복잡한 칼라 영상에서 얼굴 영역을 검색하기 위해 움직임이 있는 피부색 정보^[11]를 사용하며, 그 전체적인 구성도를 그림 1에 나타내었다. 피부 색상 움직임을 이용한 얼굴 요소 영역 검출 과정은 크게 색상변환, 색상 움직임을 이용한 얼굴 영역 검출, 얼굴 영역 내에서의 얼굴 요소 영역 검출 등의 과정으로 구성되며 다음에 그 각각의 구성에 대해 설명한다.

1. 피부 색상 분포를 이용한 색상 변환

칼라 정보를 사용하는 경우 일반적으로 사용되는 RGB 색상 공간은 휘도(Brightness) 성분에 민감하게 반응을 하므로 식 (1)과 같이 휘도 성분으로 정규화한 색상 공간을 사용한다^{[10] [11]}.

$$r = \frac{R}{R+G+B}, \quad g = \frac{G}{R+G+B}, \quad b = \frac{B}{R+G+B} \quad (1)$$

정규화된 색상 성분 r, g, b 사이에는 $r + g + b = 1$ 의 관계가 성립하므로 정규화한 색상 공간은 그림 2의 a)와 같이 r, g 평면상의 값으로 투영하여 표현할 수 있다. 정규화된 색상 성분을 사용하면 환경 및 조명에 의한 휘도 성분의 변화에 영향을 감소시킬 수 있으며, 2개의 색상 성분을 이용하므로 계산량 감소에 따른 처리가 용이해진다^{[10] [11]}.

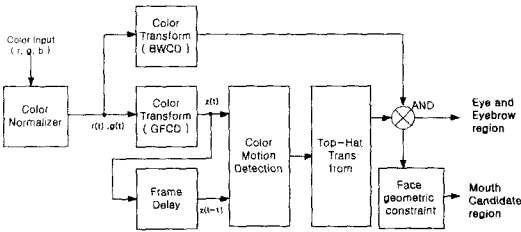


그림 1. 얼굴요소 영역 검출 알고리즘 구성도
Fig. 1. Block diagram of the facial component region extraction algorithm.

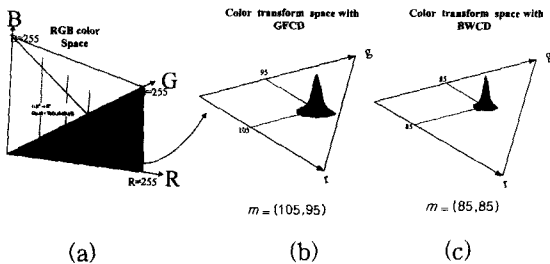


그림 2. 정규화된 색상 공간과 GFCD 및 BWCD 색상 분포 모델링 (a) 정규화된 색상 공간 (b) GFCD(Generalized Facial Color Distribution) 색상 분포 모델링 (c) BWCD(Black and White Color Distribution) 색상 분포 모델링

Fig. 2. Normalized Color space and GFCD/BWCD Modeling of color distribution. (a) Normalized color space (b) GFCD (Generalized Facial Color Distribution) Modeling (c) BWCD(Black and White Color Distribution) Modeling

정규화된 색상 공간에서 다양한 조건의 얼굴 영역에 대한 색상 발생 빈도를 조사해 보면 특정 영역을 중심으로 밀집되어 있음을 알 수 있다. 이러한 피부 색상분포는 실험에 의한 평균값 및 분산값을 이용하여 정규화된 색상 공간상에서 2차원 Gaussian 분포로 모델링 할 수 있다. 다양한 실내 조명 조건에서 실제 실

험에 의해 구해진 얼굴 피부의 평균값 과 분산값은 사람마다 어느 정도의 색상 차이가 있지만, 평균값 (m_r, m_g)은 (105, 95), 분산값은 각각 $\sigma_r=20, \sigma_g=15$ 하여 2차원 Gaussian 얼굴 색상 분포를 정하면 실내의 경우 대부분의 얼굴 색상은 포함할 수 있으며, 이를 일반 얼굴색 분포 GFCD (Generalized Facial Color Distribution)라 정의한다. 또한 얼굴영역 내의 검은 색과 흰색을 나타내는 영역의 평균값 및 분산은 실험에 의하여 (85, 85), $\sigma_r=6, \sigma_g=6$ 을 가지며, 이를 이용한 2차원 Gaussian 색상 분포를 BWCD (Black & White Color Distribution)라 정의한다. 그림 2의 b) 및 c)는 각각 정규화 된 색상 공간상에서 GFCD와 BWCD를 나타낸 것이다.

이상 설명된 색상 분포 모델을 사용하여 얼굴 색상만을 강조하는 변환을 수행할 수 있으며 이를 색상변환(Color Transform)이라 한다. 색상변환은 정규화된 색상 성분 r, g 를 그 색상 정보가 발생할 확률에 따라 밝기 값으로 변환해 주는 기능을 수행한다. 즉, GFCD를 사용하여 색상분포를 정의한 경우 그 평균값 근처에서는 높은 밝기값을 갖고, 평균값에서 멀어지는 색상을 가질수록 낮은 밝기값을 갖는다. 이를 수식으로 표현하면 다음과 같다.

$$Z(x, y) = G(r(x, y), g(x, y)) = \frac{1}{2\pi\sigma_r\sigma_g} e^{-\frac{1}{2} \left[\left(\frac{r(x, y) - m_r}{\sigma_r} \right)^2 + \left(\frac{g(x, y) - m_g}{\sigma_g} \right)^2 \right]} \quad (2)$$

여기서 $Z(x, y)$ 는 화소 위치 (x, y) 에서 색상 변환된 밝기 값이며, GFCD를 이용하는 색상변환인 경우 G 는 GFCD에서 정의된 2차원 Gaussian 함수를 나타낸다. 이러한 색상 변환은 BWCD를 이용하여서도 정의 될 수 있고 이 경우 검은 색과 흰색 영역만 강조된다. 이러한 색상 변환은 관심 있는 어떠한 색상에 대해서도 그 모델만 결정되면 적용할 수 있다. 그림 3의 b)와 c)는 각각 입력 칼라 영상3의 a)에 대해 GFCD와 BWCD를 이용하여 색상 변환한 결과를 나타낸 것이다. 각각 얼굴 살색 영역과 눈 및 눈썹 영역이 강조된 출력을 잘 나타낸다.

2. 색상 움직임을 이용한 얼굴 영역 검출

입력 칼라 영상을 색상 변환하여 얼굴 색과 유사한 색은 높은 밝기 값을 갖게 변환시켜 얼굴이 있을 것으

로 추정되는 후보 영역을 찾을 수 있으나 배경에서도 얼굴 색상과 유사한 영역이 있을 수 있다. 얼굴 영역은 배경 영역과 달리 어느 정도의 움직임이 있으므로 얼굴 색상을 가지고 있으면서 동시에 움직임이 있는 영역을 검출함으로써 얼굴 영역을 효율적으로 검출할 수 있다^[11, 12]. 본 절에서는 색상 변환 영역에서 효율적인 움직임을 검출할 수 있는 적응형 색상 움직임 방법에 대해 설명한다.

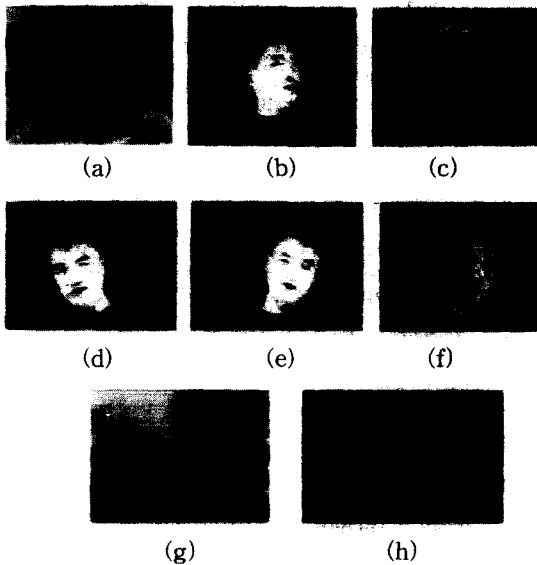


그림 3. 얼굴 성분 영역 검출 알고리즘의 결과 영상
 (a) 칼라 입력 영상 (b) GFCD 칼라 변화 영상 (c) BWCD 칼라 변화 영상 (d) 색상 변환 프레임 1 (e) 색상 변환 프레임 2 ((d) 에 움직임 이 첨가된 영상) (f) 움직임 기법을 이용한 얼굴 영역 검출 결과 (g) 배경에 살색이 있는 입력 영상 (h) (g) 영상에 대한 얼굴 성분 영역 추출 결과

Fig. 3. Results images of the facial component detection algorithm.

(a) Input color image (b) Result of the color transformation with GFCD (c) Result of the color transformation with BWCD (d) Color transformed frame 1 (e) Color transformed frame 2 (Motion is added to the (d) image) (f) Result of the detected facial region with motion color technique (g) Input image with facial color region in the background (h) Result of facial component region from image (g)

연속되는 프레임에서 색상 변환된 영상의 움직임을 측정하기 위해 식 (3)으로 정의되는 적응형 가중치 비

정합 화소 개수 측정 (Weighted Adaptive Un-matched Pixel Count :WAUPC) 방법을 사용하였다.

$$WAUPC(x, y, t) = Z(x, y, t) \otimes \sum_{i=x-N}^x \sum_{j=y-N}^y U(i, j, t)$$

$$where \ U(i, j, t) = \begin{cases} 1, & \text{if } |Z(i, j, t) - Z(i, j, t-1)| > V_{Th} \\ 0, & \text{otherwise} \end{cases}$$

$$V_{Th} = \frac{255}{1 + e^{\frac{Z(x, y, t) - 255}{2}}}$$

(3)

여기서, $Z(x, y, t)$ 는 시간 t 에서의 위치 (x, y) 에 있는 색상 변환된 화소값을 나타내고, $Z(x, y, t-1)$ 는 시간 $t-1$ 에서의 색상 변환된 화소값을 나타낸다. 또한 움직임 측정을 위한 창 크기는 $(2N+1)(2N+1)$ 이며, 두 프레임의 상응하는 창 내의 각 화소의 명암 차이가 임계값 V_{Th} 이상이면 움직임이 있는 화소 $U(i, j, t)$ 로 간주한다. 주어진 창 내에서 움직임이 있는 화소들의 개수를 구하고 이 값에 시간 t 때의 색상 변환한 영상 $Z(x, y, t)$ 를 퍼지 곱한 결과가 $WAUPC(x, y, t)$ 의 출력이 된다. 퍼지 승산 연산은 한계값 내에서는 일반 승산연산과 동일하나 그 이상에서는 포화되어 최대값을 유지한다. 임계값 V_{Th} 는 Sigmoid 함수를 이용하여 입력 영상에 적응적인 임계값을 정하게 된다. WAUPC에 의한 색상 움직임 검출 방식 결과는 얼굴 후보 영역과 움직임이 있는 영역을 동시에 공유하고 있는 영역이 강조되며, 작은 이동에 의한 겹치는 영역도 적응적으로 잘 분리하여 원하는 얼굴 영역을 검출한다^[12].

그림 3의 f)에는 이러한 방법을 사용해서 움직임이 있는 색상 변환된 영상 그림 3의 d) 및 e)에 대하여 구하여진 얼굴 영역 결과 영상을 나타내었다. 색상과 움직임을 이용한 이와 같은 방법은 사람과 같이 항상 미소한 움직임이 있는 경우 얼굴이나 손의 검출에 효율적으로 사용될 수 있으며, 다른 색상에도 색상 영역이 움직임이 있다면 조명에 별로 영향을 받지 않으면서도 복잡한 배경 하에서도 원하는 색상 영역을 영상에서 추출할 수 있는 방법이라 할 수 있다.

3. 얼굴 요소 영역 검출 과정

색상 움직임 정보를 이용해 찾은 얼굴 영역 내에서 얼굴의 특징점을 찾기 위한 전 단계로 눈과 눈썹, 입과 코 등의 얼굴 성분 후보 영역을 찾아야 한다. 먼

서 눈과 눈썹 영역을 찾기 위해 그림 1에 나타난 바와 같이 색상 움직임 추출 영역을 사용하여 검출한 얼굴 영역과, 얼굴 영역 내에서 BWCD를 이용한 색상 변환된 영상을 사용한다.

그림 3의 f)에서 보듯이 식 (3)을 이용한 색상 움직임 얼굴 영역 중 특히 눈과 눈썹 부근에서 명암 값이 작아지는 valley 부분이 많은 것을 알 수 있다. Top-hat 변환은 이러한 명암 값이 급격하게 변하게 되는 valley 영상을 추출하므로^[13] 얼굴 영역내의 이러한 valley 부분은 식 (4)에 나타난 Top-hat 변환을 사용하여 추출한다.

$$H = Z - (Z \cdot S) \quad (4)$$

$$Z \cdot S = (Z \oplus S) \ominus S$$

여기서 Z 는 색상 변환된 입력 영상이고 H 는 Hat-Top 변환된 결과 영상을 나타낸다. S 는 Morphological 연산의 seed element 이며 눈의 크기보다 작은 원형의 영역을 갖는다. 또한 \cdot , \oplus , \ominus 는 각각 Morphology 연산의 closing, dilation, 및 erosion 연산을 나타낸다. 이렇게 추출한 얼굴 영상내의 valley 영역과 BWCD 색상 변화된 영상에서 높은 명암 값을 갖는 공통 영역만을 추출하면 결과 영역이 눈과 눈썹 영역이 되며 그 결과를 그림 3의 h)에 나타내었다.

입술 영역은 위에서 구한 눈과 눈썹 영역에 기반으로 하여 얼굴 성분 요소들의 상호 위치 관계를 이용하여 입술 후보 영역을 예측한다. 예측된 입술 후보 영역 내에서 정확한 입술영역을 찾기 위하여 다시 GFCD를 이용하여 변환한 영상을 이용한다. GFCD를 이용하여 변환한 일반적인 입술 후보 영역은 피부 색상과 입술 색상의 두 가지의 색상 분포로 나눌 수 있다. 즉 색상 변환된 영상에서 높은 명암 값을 가지게 되는 피부색 색상 영역과 피부색 영역의 명암 값에 비해 낮은 명암 값을 가지는 입술 색상 영역의 두 가지 색상 영역 분포로 나눌 수 있다. 입술 후보 영역 내부의 색상 히스토그램을 구해 보면 일반적으로 두 개의 최고점을 가지게 되고 최고점 사이에서 임계점을 잡아서 이진화 하여 명암 값이 작은 값을 가지는 영역만을 추출하면 입술 후보 영역에서 원하는 입술 영역만의 영역을 추출할 수 있다. 이상의 방법으로 구한 입술영역 결과를 그림 3의 h)에 나타내었다.

Ⅲ. 자동 개인 얼굴 모델 생성

1. 모델 생성을 위한 얼굴 표현 변수의 검출

개 개인의 특징을 가장 잘 표현할 수 있는 얼굴내의 특징점은 얼굴 인식, 얼굴 모델 생성 및 얼굴 애니메이션에서 많이 연구되고 있는 부분이나, 각각의 연구에서 제시하고 있는 얼굴의 특징점은 연구의 목적과 의도에 따라 다른 위치의 특징점들을 사용하였다^{[14] [15] [16] [17] [18]}. MPEG(Moving Picture Experts Group)에서는 차세대 동영상에서 필요한 상호 대화형 동영상의 표준으로 MPEG-4를 운영 중이며, 저 전송률 전송을 통한 대화형 영상 서비스를 위해 얼굴 모델 기반 기법들을 제안하였다. 특히 MPEG-4 SNHC (Synthetic-Natural Hybrid Coding) Adhoc Group에서는 개 개인의 얼굴 모델 생성을 위해 개 개인의 얼굴을 표현할 수 있는 43개의 얼굴 표현 변수(FDP: Facial Description Parameter)를 제안하였으며^[19], 그림 4의 a)에 그 위치를 나타내었다. 본 논문에서 정의하고 사용하는 얼굴 표현 변수 특징점은 MPEG-4에서 제안하고 있는 43개 FDP 중에서 2차원 입력 영상에서 추출할 수 있는 31개의 특징점만을 선택하여 사용하며 그림 4의 b)에 본 논문에서 사용하는 특징점들을 나타내었다. 총 31개의 특징점은 각각 눈에 4*2개, 눈썹에 3*2개, 코에 5개, 입술에 4개 그리고 얼굴 윤곽선을 표현하는 8개의 점으로 이루어진다. 이러한 특징점은 얼굴 애니메이션 및 다른 얼굴 모델 연구에도 쉽게 적용할 수 있는 가장 일반적인 특징점이다.

얼굴내의 얼굴 성분 특징점 추출은 2장에서 설명한 방법으로 추출해 낸 눈, 눈썹, 코, 입 영역을 기반으로 눈, 눈썹 및 입 영역은 영역을 대표하는 색상의 분포와 경계선 정보로서 특징점을 찾아내고, 코 영역에서는 영역내의 경계선 위치와 경계선 분포 정보를 사용하여 그림 4의 b)에 정의된 특징점들을 추출한다.

그림 4의 b)에서 보는 바와 같이, 눈과 입에 관계된 특징점은 눈과 입의 모양을 타원형으로 간략화 한 경우 그 타원형의 장축과 단축이 각각 타원 경계와 만나는 화소점이 되는 것을 알 수 있다. 일반적으로 검출되는 눈 및 입 영역의 화소 분포를 구하고, 그 화소 좌표들의 eigen vector를 구하여 장축과 단축을 결정할 수 있다. 즉, 눈과 입의 후보 영역 영상을 이진화하고 화소값이 1인 화소들의 평균 좌표 값

$m(\bar{x}, \bar{y}) = \frac{1}{N} \sum_{i=1}^N p(x_i, y_i)$ 을 구한다. 이때 $p(x, y)$ 는 화소값이 1 인 화소의 위치를 나타내고 N은 화소값이 1 인 화소의 수를 나타낸다. 화소값이 1 인 화소의 평균 좌표 값 $m(\bar{x}, \bar{y})$ 을 이용하여 좌표들의 Covariance 행렬 Σ 를 구하고, Σ 의 eigen value λ 와 eigen vector ϕ 를 다음 식과 같이 구할 수 있다.¹⁾

$$\Sigma \cdot \phi = \left(\frac{1}{N} \sum_{i=1}^N p_i p_i^T - mm^T \right) \cdot \phi = \lambda \cdot \phi \quad (5)$$

이때, eigen value가 큰 eigen vector가 입과 눈 영역의 장축을 나타내고, eigen value가 작은 eigen vector 가 단축을 나타내게 된다. 그림 5의 a)는 2장에서 구해진 눈과 눈썹 영역을 나타내고, 그림 5의 b)는 구해진 눈 영역의 장축과 단축이 눈의 경계선과 만나서 결정되는 4개의 화소 점을 나타낸다. 눈썹 영역의 특징점은 그림 4의 b)에서 보인 바와 같이 눈썹의 양끝 위치와 눈썹의 가운데 상부에 위치하므로, 구해진 눈썹 영역을 둘러싸는 영역을 구하고 이의 양끝 위치와 높이 값이 가장 큰 화소의 위치를 눈썹의 특징점 후보로 선정한다.

얼굴 영역의 경계선 정보는 코의 특징점을 검출하기 위해서 뿐만 아니라, 앞에서 설명한 방법으로 추출된 눈, 눈썹, 및 입의 특징점을 보정 하기 위해서도 사용된다. 입력 명암 영상에서 경계선을 추출하기 위한 전처리 과정으로 Histogram Equalization 처리를 사용하여 명암 영상의 명암대비를 강조한 후, 3*3 Sobel 연산자를 이용하여 경계선 정보를 추출하였다. 검출된 경계선 정보는 다음과 같이 지역적으로 입력영상에 적용하는 문턱값 V_L 을 이용하여 경계선을 이진화 한다.^[21]

$$V_L = \left(\frac{1}{M*N} \sum_{i=1}^M \sum_{j=1}^N e(i, j) \right) (1+p) \quad (6)$$

여기서 $e(i, j)$ 는 (i, j) 위치에서의 경계선 값을 나타내며, 지역적으로 적용하기 위해 NxN 크기의 창을 사용한다. 식 (6)의 우변의 대괄호 안의 값은 창 내의 평균 경계선 값이 되며, 이 값에 (1 + p) 라는 가중치를 주어 적응 문턱 값으로 사용한다. 실험시 p 값은 0.6 (60%)를 사용하였다. 본 논문에서는 각각의 얼굴 성분 영역내의 경계선 정보만을 이용하므로 이진화를 위한 문턱 값 설정에 영상 전체에 적용되는 하나의 임

계값을 이용하는 방법을 사용하지 않고 각각의 얼굴 성분 영역에 따라서 적용되는 지역적 임계값을 이용하므로 이진화에 많은 경계선 정보를 잃지 않고 유지 할 수 있었다. 또한 조명이나 주변 환경에 의해 지역적으로 약해진 경계선 정보를 살릴 수 있다.

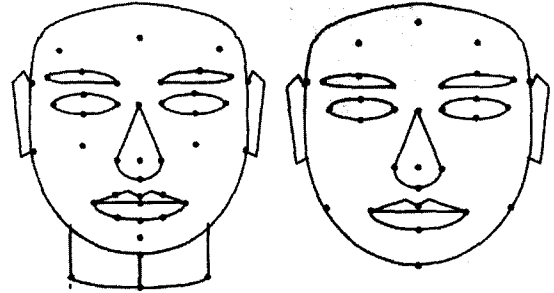


그림 4. 얼굴 표현 특징점 위치
(a) MPEG-4 SNHC의 FDP 위치 (b) 2차원 영상에서 추출 가능한 얼굴 특징점 위치

Fig. 4. Position of the facial description parameter points.
(a) Positions of the FDP proposed by MPEG-4 SNHC (b) Positions of the FDP detectable from 2D image

이상의 방법으로 구하여진 얼굴 성분 영역내의 경계선의 위치와 분포 정보를 이용하여 코 주변의 특징점 위치를 구할 수 있다. 코 주변에는 그림 4의 b)에서 보인 바와 같이 5개의 특징점이 있으며, 미간 중심점을 제외한 특징점들은 경계선 값의 수평, 수직 방향의 누적 히스토그램을 구하여 경계선의 위치와 분포 정보를 이용하여 추출한다. 그림 5의 c)는 코 후보 영역내의 추출된 경계선 영상이며, 코의 특징점 위치가 모두 수직, 수평의 경계선에 있다는 것을 알 수 있다. 따라서 그림 5의 d)와 e)에 나타난 경계선의 누적 수직 및 누적 수평 히스토그램에서 각각의 최대 값을 가지는 위치를 특징점의 위치로 결정한다. 히스토그램 분석에서 잡음에 의한 영향을 감소하기 위하여 히스토그램을 1차원 1x5 윈도우를 이용한 미디안 필터링을 취하였다.

앞서 구한 눈, 눈썹 및 입의 특징점을 부근의 경계선 정보로 보정을 한다. 이러한 보정은 눈, 눈썹 및 입의 특징점이 Erosion 과 Dilation를 이용하는 Top-hat 변환에 기초를 두고 있기 때문에 영역의 경계선 부분에서 1~3 화소 정도의 오차를 가지게 되므로 정

확성을 높이기 위하여 필요하다. Top-hat 변환의 오차 보정의 방법은 후보 특징점 주변의 24 이웃 화소를 검색하여 가장 적절한 위치에 있으며 경계선 정보가 있는 위치를 특징점으로 보정한다. 보정 전후의 특징점 위치 결과를 그림 5의 f)와 g)에 나타내었다.

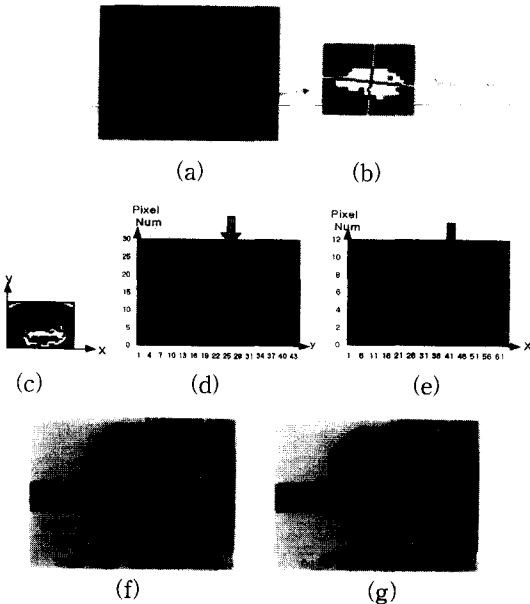


그림 5. 얼굴 요소 특징점 추출 과정

- (a) 검출된 눈과 눈썹 영역 영상 (b) 눈 영역의 장축과 단축 (c) 코 영역의 경계선 영상 (d) 코 경계선 영상의 수평 누적 히스토그램 (e) 코 경계선 영상의 수직 누적 히스토그램 (f) 특징점 보정 전 영상 (g) 특징점 보정 후 영상

Fig. 5. Extraction process of the facial component feature points.

(a) Extracted image of the eye and eyebrow region (b) Major axis and minor axis of the eye region (c) Edge image of the nose region (d) Horizontal accumulated histogram of the nose region (e) Vertical accumulated histogram of the nose region (f) Image of features before modification (g) Image of features after modification

그림 4의 b)에 나타난 얼굴 윤곽선을 나타내는 8개의 특징점들 중에서 턱에 위치한 특징점을 제외한 7개의 특징점들은 색상 변환된 영상을 이용하여 그림 6의 b)와 같이 추출된 얼굴영역을 이진화 하여 그 경계선으로부터 f)에 정의된 눈과 입의 위치 관계를 고려하여 지정된 특징점을 추출한다. 이렇게 찾아진 얼굴의

윤곽선을 나타내는 특징점 추출의 결과는 그림 6의 c)에 나타내었다. 턱에 위치한 윤곽선 특징점은 색상 변환된 이진화한 영상에선 구분이 힘들므로 그림 6의 d)와 같은 칼라 변환된 영상에서 검출된 경계선을 이용한다. 칼라 변환된 영상에서 턱의 경계선 정보가 두드러지게 되는 이유는 비슷한 얼굴 색상 정보를 가우시안 모델링을 하므로 명암차이를 확장하는 효과를 볼 수 있기 때문이다. 따라서, 그림 6의 e)와 같이 칼라 변환한 영상에서 경계선을 추출한 영상으로부터 그림 6의 f)와 같이 수평 누적 히스토그램을 구하고, 히스토그램에서 이미 알고 있는 입 영역의 아래 부분에서 가장 먼저 나오는 최대값이 나오는 위치가 입 아래의 턱 부분의 경계선이 분포되어 있는 위치임을 알 수 있다.

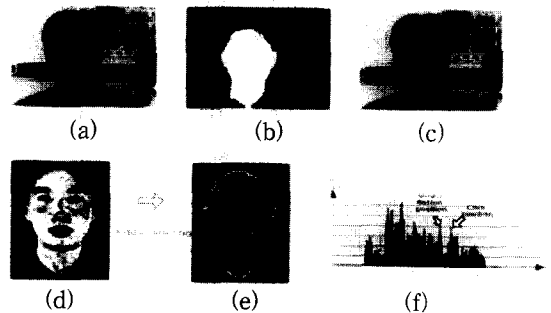


그림 6. 얼굴 윤곽선 특징점 추출 과정

- (a) 입력 영상 (b) 색상 경계 영상 (c) 얼굴 윤곽선 특징점 추출 (d) 색상 변화된 얼굴 영상 (e) 색상 변화된 얼굴 영상의 경계선 추출 영상 (f) 수평 누적 히스토그램

Fig. 6. Extraction process of the facial outline feature points.

(a) Input image (b) Color boundary image (c) Extracted feature points on the face outline (d) Color transformed face image (e) Edge image of color transformed face image (f) Horizontal accumulated histogram of face edge.

2. 일반 얼굴 모델을 이용한 개인 얼굴 변수의 정합

본 논문에서 사용하는 일반 얼굴 모델(Generic Face Model)은 그림 7의 a)에 나타난 바와 같이 1038개의 삼각형 메쉬 꼭지점 및 1703개의 삼각형 메쉬로 이루어져 있으며 사람 얼굴의 일반적인 형태를 취하고 있다. 본 논문에서 사용하는 일반 얼굴 모델은 표준화를 위해 MPEG-4의 SNHC으로 제안되고 있는 여러 연구 기관에서 제안된 모델 중 하나의 모델을

사용하였다. 2차원 영상에서 추출한 특징점을 3차원 일반 얼굴 모델에 적용하는 과정은 3차원 모델 공간과 2차원 영상 공간의 scale을 조정하는 단계와 모델의 얼굴 윤곽선 교정 단계와 교정된 얼굴 윤곽선 내에서의 눈, 눈썹, 코 및 입의 얼굴 성분 요소들의 교정 단계로서 3단계의 과정을 거치게 된다.

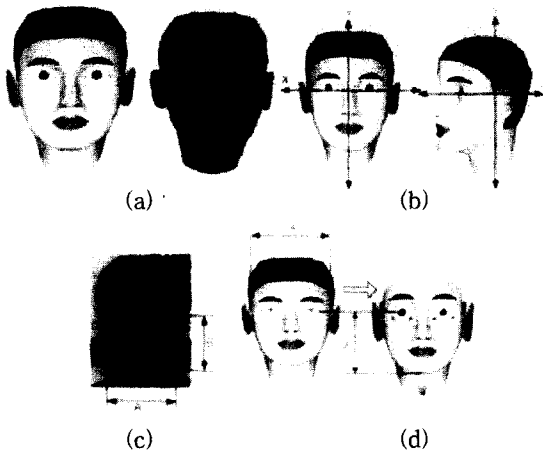


그림 7. 일반 얼굴 모델 변경
 (a) 3D 일반 얼굴 모델 (b) 일반 얼굴 모델의 기준 좌표계 (c) 일반 얼굴 모델 윤곽선 변경
 Fig. 7. Modification of the generic face model
 (a) 3D generic face model (b) Reference coordinate of the generic face model (c) Modification of generic face model outline

모델 공간과 입력 영상 공간의 scale 방법은 영상으로부터 추출된 특징점의 좌표는 2차원만의 좌표를 나타내기 때문에 얼굴의 깊이 정보가 되는 3차원 모델의 좌표에서의 깊이 값은 일반 모델에서 제공되는 값을 비례적으로 적용한다. 본 논문에서 사용하는 일반 얼굴 모델 꼭지점들의 3차원 좌표 (x, y, z)의 기준은 그림 7의 b)와 같다. 먼저 검출된 얼굴 윤곽선 특징점들을 이용하여 개인화에 적용시킬 모델의 가로 대 세로의 비가 동일하게 조절한다. 그림 7의 c) 와 같이 영상에서 검출된 특징점에 의한 얼굴의 가로와 세로 길이를 각각 A, B 라하고 이에 해당하는 모델의 길이를 각각 w, h 라 하면, 비 B/A 가 유지 되도록 3차원 모델의 길이 비 h/w 를 조절한다. 즉, 3차원 모델의 세로 길이 h는 $h = wB/A$ 가 된다. 이는 모델의 w 값을 기준으로 하여 y 값을 B/A 가 되도록 변화시키며, 모든 꼭지점에 대하여 적용한다. 개개인의 얼굴 모델의 깊이 정보는 일반 얼굴 모델의 깊이 정보

를 최대한 유지해야 하기 때문에 얼굴의 세로 비 조절시에 깊이 방향의 왜곡을 방지하기 위하여 z 축 방향으로도 같은 비로 깊이정보를 조절한다. 즉, 입의 꼭지점의 새로운 y 좌표가 y' 일 때 y 방향의 조절 거리를 Δy 이라면, 이 점에서의 z 방향의 조절거리 Δz 는 다음과 같이 주어진다.

$$y' = y + \Delta y$$

$$z : y = z' : y', \quad z' = \frac{zy'}{y} \quad (7)$$

$$\Delta z = z' - z$$

모델의 세로 비와 가로 비가 교정되면, 얼굴의 아래 부분의 윤곽선 정보라고 할 수 있는 턱 영역의 3개의 특징점의 좌표를 얼굴의 가로 세로 비를 교정하는 방법과 동일한 방법을 이용하여 얼굴의 턱 윤곽선 정보를 모델에 적용시킨다.

모델의 얼굴 윤곽에 해당하는 특징점들과 이에 해당하는 모델의 위치를 같은 방법으로 y 좌표와 z 좌표를 교정하고 사이에 있는 꼭지점들은 선형 보간을 적용하여 비례적으로 위치를 교정하여 전체적인 얼굴의 윤곽을 결정한다.

위와 같은 방법으로 모델의 얼굴 윤곽선을 교정하는 단계를 거친 후 윤곽선을 고정시킨 상태에서 얼굴 내의 얼굴 요소 교정은 추출된 얼굴 윤곽내의 눈, 눈썹, 코, 입 등 얼굴 성분들의 특징점 좌표의 위치를 적용시켜 수행한다. 얼굴 성분들의 적용 순서는 눈, 눈썹, 코, 입술 등의 순서로 적용하고, 특징점과 특징점 사이의 꼭지점들은 선형 보간법을 이용하여 변형시킨다.

IV. 실험 방법 및 결과

본 논문에서 제안한 알고리즘의 구현을 위해 사용한 입력 영상으로는 Sony Digital 캠코더 CDR-VX1000을 입력 장치로 하여 각각의 개인마다 320*240 해상도의 칼라 동영상을 60 프레임(2초)동안 받았다. 그리고 이러한 동영상 중 사람의 얼굴이 정면상을 이루는 하나의 프레임과 움직임 측정을 위한 인접 프레임을 입력 영상으로 사용하였다. 얼굴 특징점 추출 알고리즘은 Solaris 2.5 운영체제 하의 Sun Ultra Spark에서 C언어를 이용하여 구현하였으며, 추출된 특징점을 일반 모델에 적용하고 변형하여 표시하

는 데에 사용한 환경으로 윈도우 NT4.0 운영체제 하의 Pentium-200 시스템을 사용하였다. 표시에는 C를 기반으로 한 OpenGL 환경을 이용하여 생성된 개인 모델을 표시하였다.

얼굴 영역을 입력 영상으로부터 추출에 대한 실험으로서 시간 차이가 있는 두 개의 프레임 영상을 이용한 움직임 정보를 사용하는 것에 대하여 프레임간의 시간 차이가 실시간인 1/30초와 시간 간격이 어느 정도 큰 2초에 대하여 실험하였는데 두 경우 모두 모두 안정적으로 알고리즘이 적용되었다. 따라서 본 논문에서 제안하는 얼굴 영역 추출 알고리즘이 동영상에도 적용됨을 알 수 있다. 특징점 추출에 대한 실험은 5명에 대하여 입력 영상 내의 얼굴 크기와 조명 및 배경이 다르게 각각 4번씩의 실험을 하여 총 20회의 실험을 시행하였다. 제안한 칼라 영상에서의 사람의 특징점 추출에 걸린 시간은 평균적으로 약 1.7초 소요되었다. 특징점 추출의 성공 여부는 미리 수동적으로 특징점들의 위치를 정한 후 실험 후의 결과와 비교하여 특징점들의 위치가 미리 정한 특징점에서 1-2 화소의 오차 이내에 존재할 경우와 아닌 경우로서 특징점 추출의 성공과 실패 여부를 결정하였다. 총 20회의 실험에서 16회의 안정적인 특징점 추출 성공을 하여 입력 영상의 조건에 상관없이 80%의 성공 결과를 보였다. 4번의 특징점 추출 실패 중에 2회는 턱의 특징점 추출에서 실패의 결과를 보았으며 나머지 2회는 눈과 눈썹의 위치에서 실패를 하였다. 턱의 특징점 추출 실패가 다른 특징점보다 많은 이유는 턱의 특징점 추출에 이용되는 색상 변환된 얼굴 영역의 얼굴과 목의 경계선이 조명등의 환경에 따라 추출이 힘들기 때문이다. 나머지 2번의 실패인 눈썹 및 눈의 특징점 추출의 실패 경우는 머리카락이나 다른 물체에 의해 가렸을 때 안정적인 특징점 추출을 할 수 없었다. 그림 8의 결과 영상에서 알 수 있듯이 영상 내에서 개인의 얼굴이 차지하는 영역의 크기와 영상의 배경의 복잡함에 상관없이 안정적인 특징점 추출의 결과를 보였다. 본 알고리즘은 특징점 추출에 경계선 정보를 사용함으로써 안경을 쓴 경우나 얼굴내의 수염 등과 같은 인위적인 무늬가 존재할 경우들은 실험의 조건에 포함하지 않았으며 이 경우는 더욱 연구를 해야할 부분이다.

추출된 특징점을 이용하여 일반 얼굴 모델에서 개인 얼굴 모델로의 얼굴 데이터 변형에 걸린 시간은 약 0.3초 정도 소요되었다. 개개인의 2차원 얼굴 데이터

로 개인 얼굴 모델을 생성한 결과가 그림 8의 c) 및 d)에 나타내었다. 결과에서 보는 바와 같이 개개인의 얼굴의 3차원 정보인 깊이 정보는 일반 얼굴 모델에 기반으로 하고 있지만 2차원 입력 얼굴 영상을 이용하여 입력 영상내의 개인의 얼굴과 유사한 3차원 얼굴 모델을 생성됨을 알 수 있다.

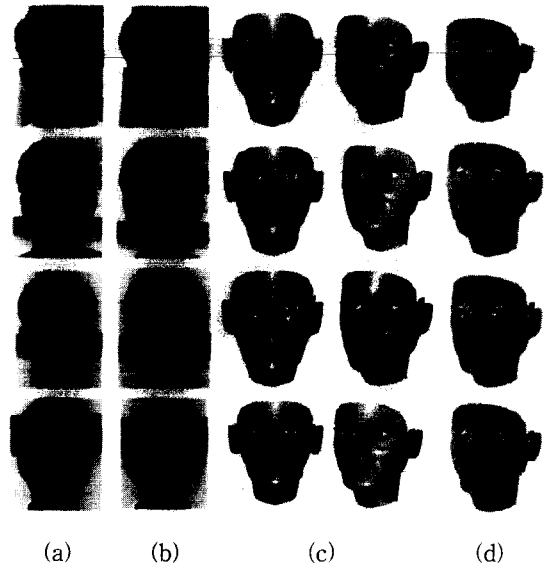


그림 8. 최종 생성된 3차원 개인 모델 결과
(a 열) 입력 영상들 (b 열) 특징점 추출 결과 영상 (c 열) 생성된 3D 개인 얼굴 모델 (d 열) 생성된 3D vertex 얼굴 모델

Fig. 8. Final results of the generated personal 3D face model.

(Column a) Input images (Column b) Result of the extracted facial feature points (Column c) Generated 3D personal face model (Column d) Generated 3D vertex face model

IV. 결 론

본 논문에서는 2차원 칼라 동영상으로부터 3차원 개인 얼굴 모델을 자동 생성하는 효율적인 방법을 다루었다. 복잡한 배경이 포함된 영상에서 얼굴 영역을 안정적으로 추출하기 위하여 피부 색상에 근거한 움직임 정보 추출 방법을 이용하였으며 실내 환경에서 잘 동작함을 보였다. 검출된 얼굴 영역 내에서 색상 정보와 경계선 정보를 활용하여, MPEG-4의 SNHC에서 제안하고 있는 개개인의 얼굴 특성을 표현하는 43개의 얼굴 특징점 파라미터(Facial Description

Parameter : FDP)중 2차원 영상에서 추출 가능한 31 개의 FDP를 자동 추출하였으며, 특별한 제약 없이 높은 추출률을 나타내었다. 추출된 2차원 얼굴 특징점들을 1038 개의 삼각형 메쉬로 이루어진 3차원 일반 얼굴 모델(Generic model)에 적용시켜 변형함으로써 개개인의 얼굴에 해당하는 모델을 자동 생성하였다. 제안하는 알고리즘은 컬러 동영상에서 배경의 복잡성, 얼굴 크기 등에 상관없이 정면상에 가까운 경우 안정적인 특징점을 추출하였으며, 팬티움 PC에서 약 2초 이내에 개개인의 얼굴 모습에 유사한 얼굴 모델을 생성할 수 있었다. 제안한 알고리즘은 배경이 복잡한 컬러 영상에서의 안정적인 얼굴 영역 검출이 가능하다는 점과, MPEG-4에서 제안된 얼굴 변수를 기반으로 하여 얼굴 특징점을 자동적으로 추출했다는 점 및 추출된 특징점을 이용하여 3D 개인 얼굴 모델을 자동 생성했다는 점 등을 특징으로 요약할 수 있다.

향후, 얼굴의 texture mapping, 귀, 머리모양 등 실제 얼굴의 모습에 유사하도록 모델을 변형시키는 방법과, 동영상으로부터 실시간으로 처리할 수 있는 방법, 그리고 얼굴 애니메이션 변수 (facial animation parameter: FAP)를 사용하여 얼굴의 표정 합성 및 문자로부터 음성을 생성할 때의 입술 모양 애니메이션 등 다양한 연구가 계속 되어야 할 것이며, 생성된 얼굴모델은 화상 회의와 등에 적용하면 극히 낮은 전송률로 대화형 화상통신이 가능할 것이다.

참 고 문 헌

- [1] Rama Chellappa, C. L. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, May, 1995.
- [2] Matthew Turk and Alex Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, vol. 12, no. 1, pp. 71-86, 1991.
- [3] MPEG-4 System Sub-Group, "MPEG-4 System Methodology and Work Plan for Scene Description", ISO/IEC JTC1/SC29/WG11 N1786, July 1997.
- [4] Nadia Magnenat Thalmann, Daniel Thalmann., "State-of-the-Art in Computer Animation" ACMCS96, Switzerland, 1996.
- [5] Yoshimitsu AOKI, Shuji HASHIMOTO, "Physical Modeling of face using spring frame based on anatomical data". *Proc. of the Multimedia Modeling Information and System*, 1997, pp. 339-354.
- [6] Won-Sook Lee, Prem Kalra, Nadia M. Thalmann "Model based face reconstruction for animation". *Proc. of the Multimedia Modeling Information and System*, 1997, pp. 323-336.
- [7] K. Aizawa, C. S. Choi, H. Harashima, T. S. Huang, "Human facial motion analysis and synthesis with application to model-based coding", *Motion Analysis and Image Sequence Processing*, Kluwer Academic publishers, 1993, pp. 317-348.
- [8] Y. Ohta, T. Kanade, and T. Sakai, "Color Information for Region Segmentation", *Computer Vision Graphics and Image Processing*, vol. 13, no. 1, pp. 222-241, May 1980.
- [9] 유태웅, 오일석, "색채 분포 정보에 기반한 얼굴 영역 추출", *정보과학회 논문지(B) 제 24권 제 2호*, pp. 180-192, 1997년 2월
- [10] S. H. Kim, N. K. Kim, S. C. Ahn, H. G. Kim, "Object Oriented Face Detection Using Range and Color Information", *Proc. of the 3'rd International Conf. on Automatic Face and Gesture Recognition (FGR'98)*, pp. 76-81, May, 1998.
- [11] 김남호, 함상진, 김형곤, 안상철, "Facial Feature Points Tracking Using Moving Color", *제10회 영상처리 및 이해에 관한 워크샵 발표논문집*, 제주, pp. 127-132, 1998년 1월
- [12] 김 남호, 김 형곤, 고성재 "색상 움직임을 이용한 얼굴 특징점 자동 추출", *전자공학회 논문집 제 35권 S편, 제8호*, pp. 55-67, 1998년 8월
- [13] R. Jain, R. Kasturi, and B.G. Schunck, *Machine Vision*, MacGraw-Hill International Edition, 1995.
- [14] A. L. Yuille, D. Cohen, P. Hallinan, "Feature Extraction From Face Using Deformable Templates", *Proceedings of*

- IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 104-109, 1989.
- [15] D. Reisfeld and H. Wolfson and Y. Yeshurun, "Detection of Interest Points Using Symmetry", Proceedings of the 3rd International Conference on Computer Vision, p. 62-65, Dec. 1990.
- [16] R. Brunelli and T. Poggio, "Face Recognition: Features Versus Templates", IEEE Trans. PAMI, vol. 15, no. 10, 1042-1052, Oct. 1993.
- [17] Litanage C. De Silva, Kyoharu Aizawa, Mitsutoshi Hatori, "Detection and Tracking of Facial Features by using edge pixel counting and deformable circular template matching", IEICE Trans. Information and System, vol. E78-D, no. 9, pp. 1195-1207, Sept. 1995.
- [18] Chung-Lin Huang, Ching-Wen Chen, "Human Facial Feature Extraction for Face Interpretation and Recognition", Pattern Recognition, vol. 25, no. 12, pp 1435-1444, 1992.
- [19] MPEG-4 SNHC Group, "Face and Body definition and animation parameter", ISO/IEC JTC1/SC29/WG11 N1365, October 1996.
- [20] K. Fukunage, Introduction to Statistical Pattern Recognition, Academic Press, New York, 1972.
- [21] I. Pitas, Digital Image Processing Algorithm, Prentice Hall International edition, 1993.

 저 자 소 개

戚尙續(正會員)

1996년 2월 연세대학교 전자공학과 졸업(공학사). 1998년 9월 연세대학교 전자공학과 대학원 졸업(석사). 1996년 9월 ~ 현재 KIST 영상미디어 연구센터 학생연구원. 주관심분야는 영상 신호처리, 얼굴 영상처리, 동영상 신호처리 등

金炯坤(正會員)

1952년 4월 5일생. 1974년 한국항공대학 항공전자공학과 졸업(공학사). 1982년 8월 Univ. of Kent(England) U.K. 전자공학과 (M.Sc). 1985년 5월 Univ. of Kent(England) U.K. 전자공학과 (Ph.D.). 1993년 7월 ~ 1994년 6월 호주 Univ. of South Australia 초빙연구원. 주관심분야는 영상처리용 VLSI 구조, 센서 fusion, 대화형 영상처리, 스테레오 비전 시스템, 가상공간 interface, 영상-그래픽스 합성, MPEG-4 등. 1977년 12월 ~ 현재 한국과학기술연구원(KIST) 책임연구원. 현재 KIST 영상미디어 연구센터 센터장