特輯

얼굴 영상을 이용한 인공현실감

崔 昌 石
明知大學校 情報通信工學科

Ⅰ. 서 론

최근에 들어 인공현실감이란 말을 자주 접하게 된다. 인공현실감(Virtual Reality)란 인간의 외계를 인식하는 시각, 청각, 촉각등의 기관에 대해서 컴퓨터에 의한 합성정보를 제시하여, 그것에 의해 인간주위에 가상적인 세계를 만들어 내기 위한 기술이다라고 정의하고 있다. 이외에도 여러가지 정의가 있지만, 컴퓨터에 의해 묘사된 가상의 세계안에서 인공적인 체험을 할 수 있도록 자유로운 상호작용과 다양한 감각기능을 가능하도록 하는 내용을 포함하고 있다.[1][2]

인공현실감의 기술중에서 가장 중요하다고 생각되는 것은 臨場感, 對話性, 自律性의 3가지 요소이다. 임장감이란 자신의 눈앞에 현실이라고 착각할 정도의 현실감있는 정보를 제시하는 것이다. 대화성은 인간으로부터 가상세계에 정보입력이 이루어져, 그에 대한 응답이 실시간으로 이루어지는 것이다. 최후의 요소인 자율성이란 던져진 물체가 포물선을 그리며 낙하하여 바닥에 충돌한 뒤 튀어오르듯이, 미리정해진 법칙에 따라 자율적으로 운동하는 것이다. 현재까지 이러한 요소를 모두 만족하고 있는 시스템은 찾아보기 힘들다. 고도의 인공현실감을 추구하기 위해서는 시각디스플레이, 청각디스플레이, 촉각디스플레이등 장치기술도 발달되어야 하지만, 표현하려는 정보를 현실감있게 가시화하는 기술도 매우 중요하다. 본고에서는 인간의 신체부위 중에서 정보를 가장 많이 입출력하는 얼굴에 주목하여, 얼굴에 나타난 여러가지 정보를 얼굴영상으로 가시화 하여 인공현실감을 추구하는 연구에 대해서 살펴보고져 한다.

이하 II장에서는 얼굴에 나타난 정보를 살펴보고, 얼굴의 정보를 3차원적으로 분석·합성하기 위한 얼굴의 3차원 모델을 구성하는 방법을 소개한다. III장과 IV장에서는 얼굴표정 정보를 합성하고 분석하는 방법을 소개한다. V장에서는 얼굴의 특징을 나타내는 정보와 나이정보에 주목해서 얼굴영상을 합성하는 방법을 서술한다. VI장에서는 이와같이 합성된 얼굴영상의 응용에 대해서 살펴보고자 한다.

Ⅱ. 얼굴영상의 3차원 처리의 개념

얼굴영상의 3차원 처리에 있어서 기본개념은 그림 1과 같다. 즉, 얼굴영상에 얼굴의 정보를 부가하여 새로운 얼굴영상을 생성하는 것을『합성』이라하고, 반대로 얼굴영상으로부터 얼굴의 정보를 추출하는 것을『분석』이라고 한다. 예를들면, 합성은 어떤 개인의 사진에『웃는다』정보를 부가하면, 웃는 얼굴을 생성하는 것을 말하며, 분석은 웃는 얼굴사진으로 부터『웃는다』라는 정보를 추출하는 것을 말한다.

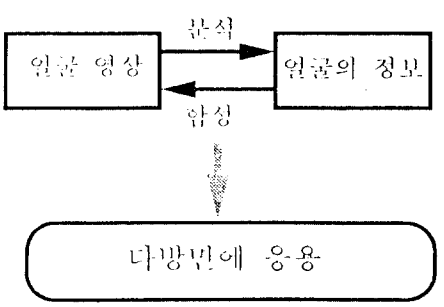


그림 1. 얼굴영상의 3차원처리의 개념

얼굴영상의 3차원 처리는 비교적 개별적으로 발달되어 왔던 컴퓨터 그래픽스 기술, 컴퓨터비젼 기술, 영상처리기술을 복합적으로 사용하여 현실감있는 고품질의 영상을 얻을 수 있을 뿐만아니라, 인간의 자연스런 3차원 동작패턴도 용이하게 획득할 수 있는 새로운 기술이며, 응용분야가 넓은 것이 특징이다.

1. 얼굴의 정보

인간이 사회생활을 영위하는데 있어서 얼굴은 다양한 기능을 가지고 있다.[1-5] 예를들면, 얼굴에는 개인의 특징을 나타내는 기능, 개인상호간의 대화기능, 시각, 청각, 촉각, 미각등의 감각기능, 음식물의 섭취기능 등이 있다. 이중에서 개인의 특징을 나타내는 기능이나 대화기능은 다음의 3가지 신호에 의해 이루어진다고 생각되고 있다.

즉, 얼굴의 형상, 골격, 피부색, 얼굴 각부위(눈썹, 코, 입술 등)의 위치관계와 같이 거의 변하지 않는 것, 주름과 같이 천천히 변하는 것, 입술의 움직임과 같이 빨리 변하는 것이 있다. 변하지 않거나 천천히 변하는 신호로 부터는 인종·성별·연령·성격·성품·개성·건강상태 등의 정보를 파악할 수 있고, 빨리 변하는 신호는 개인상호간의 대화에 있어서 중요한 역할을 하고 있다. 표정은 감정·정서·심리상태를 나타내고, 입술과 혀의 움직임은 발음의 애매함을 줄여주고, 頭部의 움직임은 情動表出이나 의지전달에 관련된 기능을 갖고 있다. 다시말하면, 얼굴을 한번 보면, 개인에 대한 대부분의 특징을 알 수 있다는 의미이다. 또, 눈으로 보고, 입으로 말하고, 귀로 듣고, 코로 냄새 맡는 등의 기능을 감안해 보면, 인간의 정보 입·출력의 대부분이 얼굴에서 이루어지고 있음을 알 수 있다. 이러한 이유로 인공현실감을 위한 입·출력장치도 많은 부분이 얼굴과 관련이 있다고 생각된다.

2. 얼굴의 3차원모델의 구성

얼굴영상을 3차원 얼굴의 2차원 화면상에 투영이라는 관점에서 볼 때, 얼굴영상의 분석과 합성은 3차원적으로 이루어져야 한다. 2차원 얼굴영상을 3차원적으로 분석과 합성을 하기 위해서는, 영상속의 얼굴에 대한 충실한 3차원 모델이 필요하다. 여기에서는, 얼굴의 일반적인 3차원형상을 표현하기 위해서, 얼굴을 점과 선으로 근사한 얼굴의 3차원 형상모델을 준비한다.[6] 이것을 얼굴의 표준 형상모델로 한다.

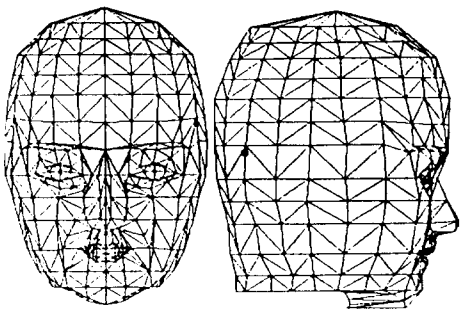


그림 2. 얼굴의 3차원 형상모델

그림 3. 대상인물의 정면상

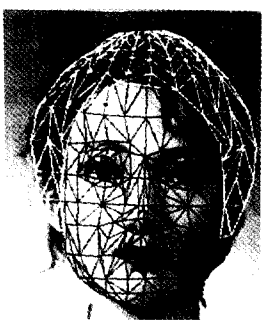


그림 4. 정합된 얼굴의 형상모델

이것은 약 600개의 꼭지점과 약 700개의 삼각형패치(patch)로 이루어져 있다. 이 모델을 대상인물의 정면상(그림3)에 정합하여, 개인 얼굴의 3차원 형상모델을 얻는다. 이때, 표준형상모델을 3차원 어핀변환(affine transform)을 통하여 얼굴의 크기, 위치 및 방향을 조절한다. 또한, 얼굴의 윤곽 및 각부위(예를들면, 눈썹, 눈, 코, 입 등)는 가능한한 정밀하게 정합한다.(그림4)

이어서, 개인얼굴의 형상모델의 삼각형 패치에 대해서 얼굴의 표면을 정의하여, 이것에 정면상의 휘도

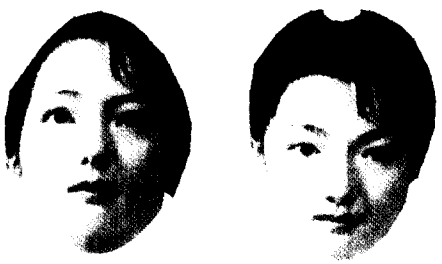
정보를 투영함으로써 대상인물에 대한 얼굴의 3차원 모델을 구성한다. 이와같은 얼굴의 3차원 모델의 구성은 2차원의 얼굴영상으로부터 얼굴의 3차원구조를 복원하는 작업이므로, 일단 모델이 구성되면 상하좌우 여러 시점으로 부터 본 얼굴영상을 합성할 수 있다.(그림5) 1매의 얼굴영상으로부터 얼굴의 3차원 구조를 정확히 복원하는 작업은 원리상 불가능한 것이지만, 인간의 일반적인 얼굴의 3차원 형상으로 부터 대체적으로 추정함으로써, 얼굴의 정보를 가시화하는 데에 있어서는 유효하다고 생각 된다. 그림 5는 두부의 3차원 동작을 현실감있게 가시화 해본 것이다.

Ⅲ. 얼굴표정의 합성

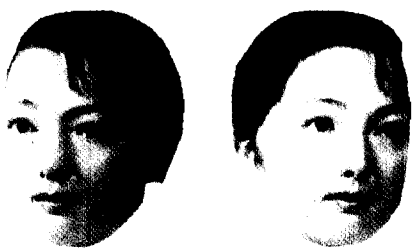
1. 얼굴의 형상모델의 변형규칙

얼굴표정은 얼굴의 3차원모델을 변형하여 컴퓨터내에서 합성한다. 문제는 얼굴의 3차원모델을 변형하는 방법이다. 왜냐하면, 얼굴의 골격구조, 얼굴근육의 움직임에 따라 표정이 변화하기 때문에, 인간이 지을 수 있는 표정이 있는가 하면 지을 수 없는 표정도 있다. 예를들면, 양눈썹사이(미간)는 좁아지기는 하지만, 넓어지는 경우는 없다. 이것은 미간을 좁게하는 근육은 있어도 넓게하는 근육은 없기 때문이다. 심리학분야에서는 Ekman과 Friesen에 의해 제안된 FACS(Facial Action Coding System)라는 표정기술법이 널리 알려져 있다.[4] FACS에서는 얼굴근육의 위치와 움직임 방향을 해부학적으로 고려하여 얼굴근육의 수축·이완에 따른 얼굴표정의 변화를 AU(Action Unit)라는 44개의 기본동작으로 분해하고 있다. 역으로, 모든 표정은 44개의 AU의 조합으로 기술될 수 있다고 한다.

그림 6은 얼굴근육의 일부를 나타내고 있다. 예를들면 ①번 근육이 수축하면, 눈썹내측이 올라가게 된다(AU1). 이와같이 얼굴표정이 얼굴근육의 수축과 이완에 따라 변한다는 것에 착안하여 얼굴의 3차원모델을 변형하면, 얼굴표정을 자연스럽게 합성할 수 있다.



(a) 상 (b) 하



(c) 좌 (d) 우

그림 5. 상하좌우 여러시점에서 본 얼굴의 합성

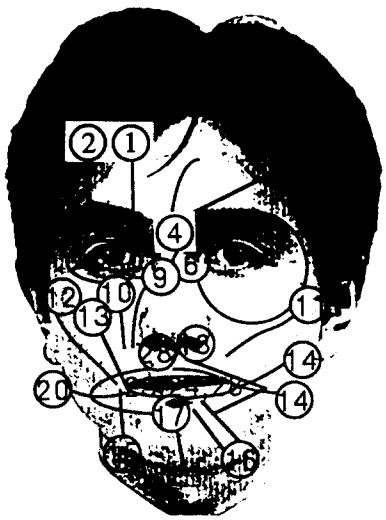


그림 6. 얼굴의 근육

그림 7은 눈과 눈썹에 대한 표정변화를 컴퓨터내에서 실현하는 방법을 설명한다.[6][8][9][12] AU1은 눈썹이 올라가는 동작이므로, 무표정(점선)에서 그림7의 AU1과 같이 얼굴의 형상모델이 변형되도록 규칙을 정해놓는다.(실선) 이와같은 방법을 입, 코 등에도 적용하면 얼굴표정을 변하게 할 수 있다. 이와같이, AU의 변형규칙을 실현하면, AU의 번호와 강도를 지정하므로서 그에 대응하는 얼굴표정을 용이하게 합성할 수 있을 뿐만아니라, AU의 조합으로 다양한 얼굴표정을 자유자재로 합성할

수도 있게 된다. 나아가서, AU의 조합을 연속적으로 변화시키면 얼굴동영상을 합성할 수 있게 된다.

표 1. 안면근육의 기본동작의 일람

AU no.	AU 명	AU no.	AU 명
1	눈썹 내측을 올린다	20	입술양단을 옆으로 끈다
2	눈썹 외측을 올린다	23	입술을 강하게 다문다
4	눈썹을 내린다	24	입술을 상하로 누른다
5	윗눈꺼풀을 올린다	25	턱을 내리지 않고 아래입술을 내린다
6	뺨을 올린다	26	턱을 내리면서 아래입술을 내린다
7	눈꺼풀을 긴장시킨다	27	입을 크게 벌린다
8	입술을 서로 접근시킨다	28	입술을 빨아들인다
9	코에 주름을 잡는다	29	아래턱을 내민다
10	윗입술을 올린다	30	턱을 좌우로 이동시킨다
11	콧볼홈을 깊게 한다	32	입술을 깨문다
12	입술양단을 끌어 올린다	35	볼을 빨아들인다
13	입술양단을 예리하게 끌어올린다	41	윗눈꺼풀을 내린다
14	보조개를 만든다	42	눈을 가늘게 뜬다
15	입술양단을 내린다	43	눈을 감는다
16	아랫입술을 내린다	44	눈을 작게 뜬다
17	턱을 올린다	45	눈을 깜빡인다
18	입술을 좁힌다	46	윙크
			안구의 회전

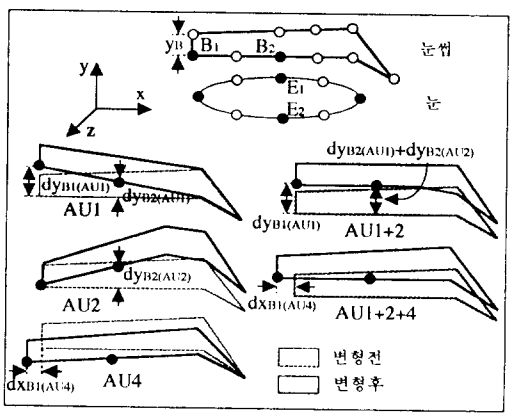


그림 7. 얼굴근육에 따른 형상모델의 변형규칙

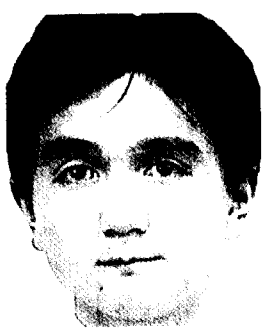


그림 8. 무표정의 얼굴

2. 표정의 합성예

현재, 34개의 AU를 컴퓨터내에서 실현하고 있다. 그 일람은 표1과 같다. 표정의 합성은, 무표정의 형상모델의 눈썹, 눈, 코, 입부분을 상기의 방법에 따라 AU의 강도만큼 움직이고, 무표정의 얼굴영상의 휘도치를 삼각형마다 텍스쳐 매핑하여 행한다. 무표정의 얼굴영상을 그림 8에, 각 AU의 예를 그림 9에, AU의

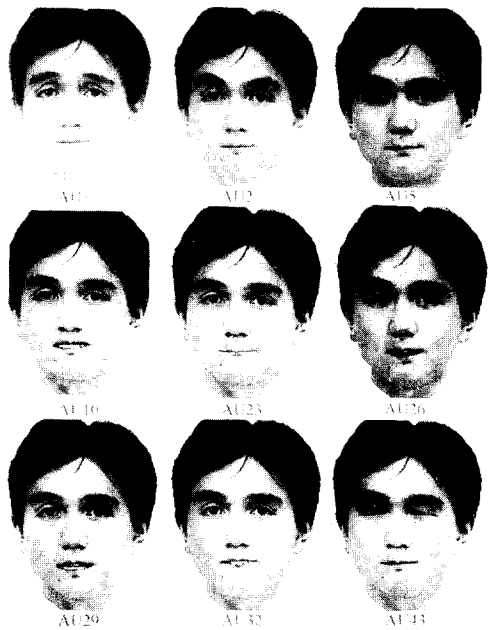


그림 9. AU에 대한 합성표정

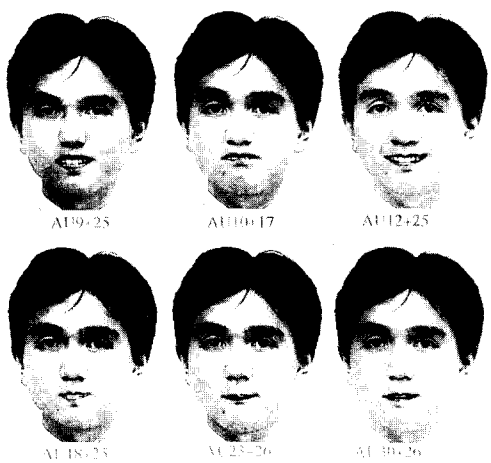


그림 10. AU조합에 대한 합성영상

조합의 합성 예를 그림 10에 나타낸다. 또, 기본표정(놀람, 공포, 혐오, 화냄, 기쁨, 슬픔)합성예를 그림 11에 나타낸다. 또, 하나의 예로서, 모나리자의 표정을 합성했다. 모나리자의 합성영상을 그림 12에 나타낸다. 이들의 인물에 대한 합성 예에서는 표1을 참고하면, AU의 의미가 잘 나타나고 있고, 실제의 사진과 같이 현실감있는 얼굴영상이 얻어진 것을 알 수 있다.

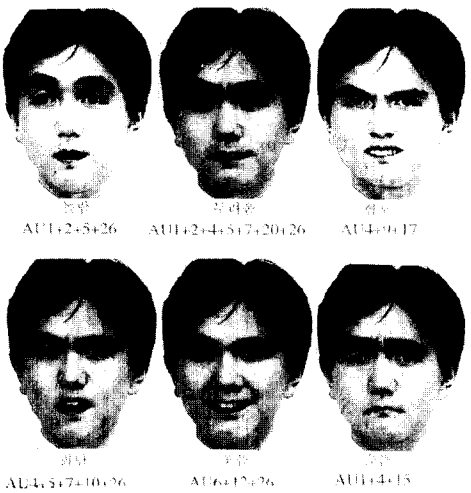


그림 11. 희노애락에 대한 합성표정

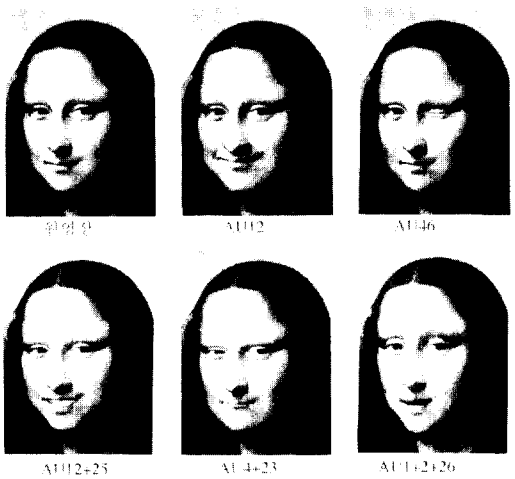


그림 12. 모나리자의 합성표정

3. 한글에 부합한 입모양 합성

이와같이 여러가지 표정을 합성할 수 있다면, 같은 방법으로 입모양도 자유자재로 변형할 수가 있게 된다. 한글은 모음이 중모음을 포함하여 19개, 자음이 중자음을 포함하여 21개로서, 이들의 조합으로 생성될 수 있는 문자는 총 14,364자에 이른다. 그러나, 입모양을 대개 모음에 의존한다는 것에 착안하면, 모음발음에 필요한 입모양 8개(ㅏ, ㅓ, ㅗ, ㅜ, ㅣ, ㅐ, ㅔ), 자음발음에 필요한 입모양 2개(입술소리, 그외소리)를 기본형으로 분류할 수 있다.[13]

중모음의 경우는, 기본형 모음의 연속적인 변화로 표현할 수 있기 때문에, 10개의 입모양 패턴을 연속적으로 연결함으로써, 우리말을 표현하는 동영상을 합성할 수 가 있다. 그림 13이 분류된 기본형 입모양을 합성한 것이다.

그림 13. 한글발음에 필요한 입모양의 기본형

Ⅳ. 얼굴표정의 분석

III장에서 서술한 방법으로 정지영상 하나하나를 현실감 있게 합성하였더라도 이들을 연결하여 자연스러운 동작을 갖는 동영상을 생성하는 것은 또 하나의 문제이다. 이 경우, 인간이 실제로 움직이는 얼굴동영상으로부터 두부의 움직임, 표정의 변화과정, 연음절 발음과정에 대한 전형적인 패턴을 분석해서, 각종

패턴의 데이터베이스를 사전에 구축해 놓고, 이들 패턴을 편집, 가공, 조합하면, 자연스런 얼굴동영상을 생성할 수 있을 것으로 기대된다. 이들 패턴에 있어서는 두부의 움직임은 3차원 동작파라미터, 표정변화와 입모양 변화는 얼굴근육의 움직임, 즉 AU로 기술되기 때문에, 한번 추출된 패턴은 대상인물이 바뀌어도 그대로 적용될 수 있다.

1. 얼굴의 3차원 움직임의 표현

얼굴과 같이 표정이 변화하면서 두부가 3차원으로 움직이는 비강체의 운동은 일반적으로 다음과 같이 기술할 수가 있다.

$$\text{Trans}(\text{Rot}(\text{Deform}(P))) \quad (1)$$

얼굴의 경우는 표정이 변화한 후, 두부가 3차원 운동(회전운동과 병진운동)을 하는 것으로 해석될 수 있다. 식(1)을 정식화 하면,

$$P' = R(P + \delta P) + T \quad (2)$$

가 된다. 여기서 P는 표정에 의한 비강체 움직임이고, P와 P'는 각각 운동전후의 점의 3차원 위치벡터이다. R과 T는 각각 회전운동과 병진운동을 나타낸다.텔리비젼 카메라로 찍은 동영상을 해석하는 경우, 각 프레임간 (1/30 초)의 움직임은 작다고 가정할 수 있다. 이 경우, 회전각이 작다고 가정하여 $\cos r_x \simeq 1$, $\sin r_x \simeq r_x$와 같이 1차 근사를 이용하면, 식(2)는 다음과 같이 구체적으로 기술할 수 있다.[14]

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & r_z & -r_y \\ -r_z & 1 & r_x \\ r_y & -r_x & 1 \end{bmatrix} \begin{bmatrix} x+\delta x \\ y+\delta y \\ z+\delta z \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} \quad (3)$$

여기서 $P' = (x', y', z')^T$, $P = (x, y, z)^T$, $\delta P = (\delta x, \delta y, \delta z)^T$, $T = (t_x, t_y, t_z)^T$이고,

$$R = \begin{bmatrix} 1 & r_z & -r_y \\ -r_z & 1 & r_x \\ r_y & -r_x & 1 \end{bmatrix}$$

이다. r_x, r_y, r_z는 각각 x, y, z축을 중심으로한 회전각이고, tx, ty, tz는 각각 x, y, z축 방향의 병진운동량이다. 식(3)에서 회전운동은 각속도, 병진운동은 속도를 이용하며, 정리하면 다음과 같이된다.

$$\begin{bmatrix} v_x \\ v_y \\ v_z \end{bmatrix} = \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} \begin{bmatrix} x+\delta x \\ y+\delta y \\ z+\delta z \end{bmatrix} + \begin{bmatrix} u_x \\ u_y \\ u_z \end{bmatrix} \quad (4)$$

이 식은 벡터를 이용해서 다음과 같이 표기할 수 있다.

$$V = \Omega(P + \delta P) + U \quad (5)$$

여기에서, $V = (v_x, v_y, v_z)^T$, $U = (u_x, u_y, u_z)^T$이고,

$$\Omega = \begin{bmatrix} 0 & \omega_z & -\omega_y \\ -\omega_z & 0 & \omega_x \\ \omega_y & -\omega_x & 0 \end{bmatrix} \quad (6)$$

이다. 이 식을 다시 정리하면

$$V_h \equiv \Omega P + U \quad (7)$$

$$V_f \equiv \Omega \delta P \quad (8)$$

$$V = V_h + V_f \quad (9)$$

이 된다. 여기에서 V_h와 V_f가 각각 두부의 운동과 안면의 움직임을 나타내는 속도벡터이다. 두부의 움직임에 비해 안면의 움직임이 적다 (즉, $|Vh| \gg |Vf|$)고 가정하여, 두부의 3차원 움직임을 먼저 추정한 후 안면표정을 추정하기로 한다.[11][12]

2. 두부의 3차원 운동정보의 추정

연속하는 2매의 영상간에서 두부의 파라미터를 추정한다. 먼저, 영상에서 화소의 휘도의 변화와 속도간에는 다음과 같은 관계가 있다.[15]

$$I_x v_x + I_y v_y + I_t = 0 \quad (10)$$

I_x, I_y, I_t는 각각 x, y, t 축 방향의 휘도치의 편미분이다. 식(9)에서 일단 V_f를 일단 무시하면 두부를 강체로 취급할 수 있다. 식(9)의 V_x, V_y를 식(10)에 대입하면

$$(-zI_y, zI_x, xI_y - yI_x, I_x, I_y)(\omega_x, \omega_y, \omega z, u_x, u_y)^T = -I_t \quad (11)$$

와 같이 된다. 이식을 각화소에 대해서 적용하면 추정식이 선형으로 유도되어 회전운동과 병진운동을 최소자승법으로 구할 수 있다. 여기에서 구한 두부의 3

차원 운동파라미터는 실제값에서 약간 벗어났다고 볼 수 있다. 왜냐하면, 안면표정의 움직임인 비강체운동을 무시했기 때문이다. 비강체 운동을 한 화소를 제외하고 식(11)로 부터 재추정하면 좀더 정확한 값을 얻을 수 있다(문헌[11]참조). 이 방법은 특징점이나 윤곽선 등을 추출하지 않고, 영상으로부터 두부의 3차원 파라미터를 직접 구할 수 있는 것이 특징이다

3. 얼굴표정 정보변화의 추정

전절에서와 같이 두부의 3차원 운동파라미터를 추정하면, 그 파라미터를 이용하여 두부의 3차원 동작을 합성할 수 있다. 합성된 영상과 그다음 프레임 간에 변화는 표정에 의한 것이다. 먼저, AU의 변형규칙으로부터 AU의 강도와 얼굴의 형상모델의 절점의 속도와의 관계를 식(12)와 같이 정식화 할 수 있다.[11]

$$V_g = V_A a + \epsilon \quad (12)$$

$$V_A = (V_{A1}, V_{A2}, \cdots, V_{AN}) \quad (13)$$

단, $a=(a_1, a_2, \cdots, a_n)^T$는 AU의 강도를 나타내는 벡터이고, N은 AU의 수이다. 또, V_A는 AU의 변형규칙을 기술하는 행렬이고, 벡터 $V_{Aj} = (v_{Axij}, v_{Ayij}, v_{Azij}, \cdots, v_{AzMj})^T$, $(i \le i, j \le m)$는 제 j번째 AU의 변형에 대한 속도벡터를 나타낸다. 그 요소 v_{Axij}는 제 j번째 AU의 단위 강도에 대한 제 i번째 절점의 x방향의 속도이다. 식(12)와 (13)을 식(11)에 대입하면 화소의 속도와 AU의 강도의 관계를 선형식으로 유도하여, 최소자승법으로 AU의강도

(a) 원영상 (b) 형상모델 (c) 합성영상

그림 14. 얼굴동영상의 분석과정

a를 추정할 수 있다. 그림 14가 CCITT표준영상 Claire의 동영상으로부터 얼굴표정을 분석하는 과정을 나타낸 것이다. 제 1프레임은 II.2절의 방법으로 얼굴의 형상모델을 정합한다. 제 2프레임 부터는 본 방법을 이용하여, 두부의 3차원 파라미터와 AU의 강도를 추정한다. 그림 14(b)는 제 8프레임까지 추정하여, 추정된 파라미터를 이용하여 합성된 얼굴형상모델을 원영상과 비교한 것이다. 그림14(c)는 제 1프레임의 영상을 이용하여 제 8프레임을 합성한 영상이다. 그림 15는 제 3, 제 5, 제 8프레임을 분석하여 합성한 영상과 원영상을 비교한 것이다. 이 결과로 부터 두부의 3차원 동작, 얼굴표정이 잘 재현되어 있으므로, 표정분석이 올바르게 이루어져 있음을 알 수 있다. 이와같은 표정분석방법을 이용하면, 얼굴동영상으로부터 인간의 자연스런 얼굴의 3차원 동작패턴을 추출할 수 있다.

(a) 원영상

(b) 합성영상

Frame #3 Frame #5 Frame #8

그림 15. 원영상과 합성영상의 비교

V. 얼굴 특징정보의 합성

1. 얼굴에 대한 가설

얼굴에는 표정에 관한 정보만이 있는 것이 아니라, II장에서 언급한 바와 같이 인종, 연령, 성격, 성품, 개성, 건강상태등 개인의 특징을 나타내는 여러가지 정보가 있다. 이와같은 정보를 합성하기 위해서 필자

는 얼굴영상에 대하여 그림 16과 같은 가설을 세우고 있다.[16] 즉, 한사람의 얼굴은 몇 사람의 얼굴의 합성이라는 것이다. 이 가설을 설명하기 위해 한가지 예를 들어본다. 우리는 어린애의 얼굴에서 엄마를 닮았나, 아빠를 닮았나를 찾는다. 엄마를 많이 닮은 경우도 있고, 아빠를 많이 닮은 경우도 있다. 경우에 따라서는 할아버지, 할머니를 닮은 경우도 있다. 이것을 영상합성의 입장에서 바꾸어서 말하면, 어린애의 얼굴은 엄마, 아빠, 할아버지, 할머니의 얼굴을 기저로 해서 합성된 얼굴이라는 의미이다. 또한, 같은 부모에서 태어난 형제도 얼굴이 닮아 있으면서도 각양각색으로 다른 것은 닮아 있는 정도의 비가 다르기 때문일 것이다. 옛날 교통수단이 불편했던 점을 감안하면, 이웃마을끼리 결혼이 성했기 때문에 얼굴에 지방적 특성이 있었음직 하다. 시야를 넓혀서 생각해보면, 우리나라 인구가 남북한을 합하면 7천만이라 하지만, 얼굴 인자를 분석해보면 그리많지 않은 수로 우리민족의 얼굴을 모두 합성할 수 있지 않을까 하고 생각하고 있다. 이러한 인자를 체계적으로 분석·정리함으로써 우리민족의 기원을 더듬을 수도 있을런지도 모른다.

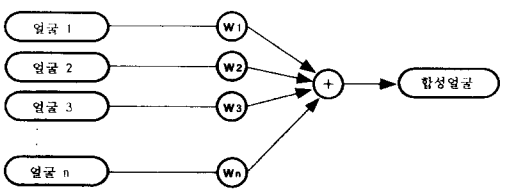


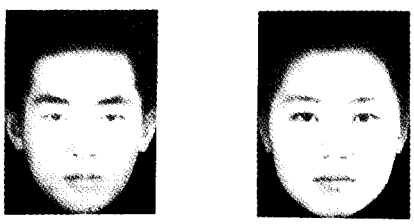
그림 16. 얼굴특징정보의 합성을 위한 가설

2. 얼굴의 형상과 텍스쳐의 합성

이와같은 가설하에서 얼굴의 특징정보를 합성하기위해, 기저얼굴에 얼굴형상모델을 그림 4와 같이 정합한 후, 얼굴형상모델의 각절점을 다음과 같이 내분한다.[16]

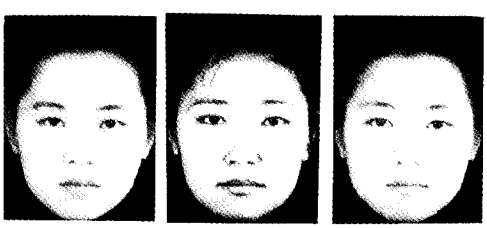
$$P' = \Sigma\ W_i P_i \qquad (14)$$

여기서, P와 P′는 각각 기저얼굴과 합성후의 형상모델의 절점의 3차원 좌표이고, W_i는 각 기저얼굴에 대한 닮음비 이다. 또한, 각화소에 대해서도 같은 방법으로 내분하여 합성영상을 얻을 수 있다. 그림 17은 20대 남여의 평균얼굴을 합성한 것으로, 20대 얼굴을 대표한 것이라고 생각할 수 있다. 우리나라 남자는 평균치에 가까운 사진5매를 합성한 것이고, 여자는 50매를 합성한 것이다. 그림 18은 우리나라 각 지방의 평균얼굴을 그려본 것이다. 그림 19는 일본, 중국, 태국의 20대 여성의 평균얼굴을 가시화 한 것이다. 나아가서, 이 방법을 확장하여 눈, 눈썹, 코, 입 등 얼굴 각 부위별로 각각 다른 비로 내분함으로써 좀더 다양한 얼굴영상을 얻을 수 있다. 그림 20은 눈썹, 눈, 코, 입의 각 부위를 바꾸어 본 것이다. 그림 20(a)는 한국여자의 코를 일본여자로 바꾸었고, 그림 20(b)는 중국여자의 눈과 눈썹을 일본여자의 것으로, 그림 20(c)는 입을 태국여자의 것으로 바꾸어서 합성한 것이다. 이와같이 하면, 가상의 인물을 현실감있게 고품질로 가시화할 수 있어, 지방 또는 나라에 따른 얼굴의 특징을 비교·검토할 수도 있다.

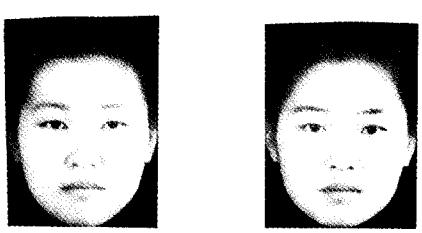


(a) 남자 (b) 여자

그림 17. 우리나라 20대 남여의 평균 얼굴



(a) 경기도 (b) 강원도 (c) 충청도

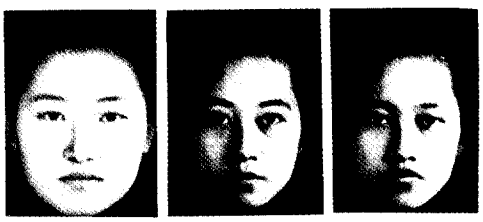


(c) 경상도 (d) 전라도

그림 18. 우리나라 각지방의 20대 여성의 평균얼굴

(a) 일본 (b) 중국 (c) 태국

그림 19. 일본, 중국, 태국의 20대 여성의 평균얼굴



(a) (b) (c)

그림 20. 평균얼굴의 각부위별 변화

3. 나이정보의 합성

얼굴을 보면 나이를 대개 짐작하게 된다. 이것은 얼굴에는 나이를 짐작케하는 성분이 있다는 의미이다. 이성분이 많으면 나이가 들었다고 생각하고, 이성분이 적으면 나이가 덜 들었다고 생각하는 것은 아닐까? 그렇다면, 나이성분을 추출하게 되면, 한장의 사진만 있어도 그사람의 일생 동안 변하는 모습을 그려볼 수 있지 않을까 생각된다. 나이변화를 위하여, 10대부터 60대까지 연령별로 얼굴영상을 골고루 수집하여 주성분분석을 이용하여 나이성분을 추출한다.[17][19]추출된 나이성분을 남·여 평균영상에 선형적으로 가감함으로써 나이를 변화한 영상을 그림 21에 나타낸다. 이 그림으로 부터, 나이변화 과정을 잘 알 수 있다.

Ⅵ. 응 용

얼굴영상의 분석·합성·처리에 대한 연구는 컴퓨터를 이용하여 얼굴의 정보를 표현하고, 분석하는 기술로서 다방면에 응용을 창출할 수 있다. 이 기술은 차세대 영상통신, 휴먼인터페이스, 이산가족찾기, 방송, 영화, 예술, 광고, 컴퓨터를 이용한 교육(CAE), 범죄수사, 의학, 심리학, 무인안내, 114안내 등에 응용이 가능하여, 과학기술뿐만아니라 학문 및 사회전반에 걸쳐 광범위하게 응용이 될 것으로 기대된다. 이하, 구체적인 응용에 대해서 간단히 정리한다.

1. 차세대 영상통신

현재, 영상통신에 있어서는 화소간의 통계적인 성질을 이용한 영상압축방식(MC-DCT, DPCM, 변환부호화, VQ, Subband 부호화 등)이 주류를 이루고 있다. 얼굴영상의 분석합성법을 이용하면, 두부의 3차원 동작파라미터와 얼굴근육의 움직임만을 송수신하는 지극히 효율적인 통신을 실현할 수 있어 공중전화망을 통한 영상전송이 가능하다. 또한, 현재의

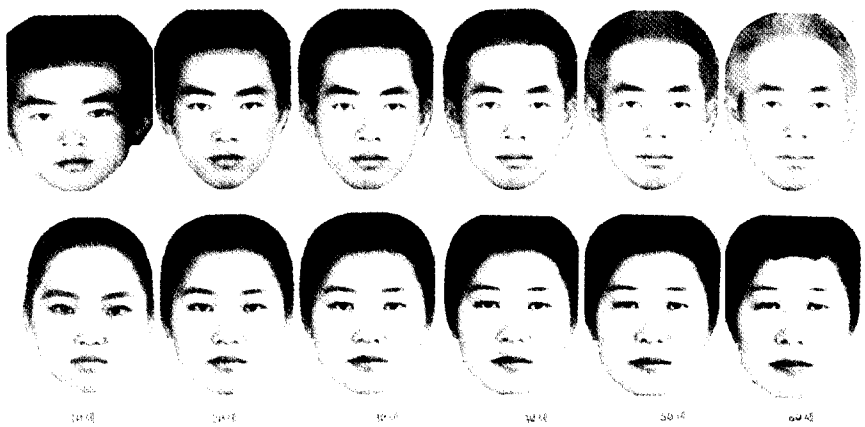


그림 21. 남여 평균얼굴의 연령변화

방식으로 실현곤란한 고부가가치통신의 실현도 용이하다. 이러한 통신방식을 모델기반 부호화(Model-Based Coding)이라고 하여 차세대 영상통신방식으로서 ISO/IEC MPEC IV에서 표준화 검토중이다. 모델기반 부호화에 있어서 얼굴영상의 분석합성법이 핵심기술이 된다.

2. 휴먼인터페이스

얼굴표정과 음성을 이용한 인간의 풍부한 표현력을 각종 전자제품이나 기계에 부여하므로서, 인간중심의 휴먼인터페이스를 실현할 수 있다. 휴먼인터페이스가 실현되면, 인간과 기계와의 대화가 인간상호간의 대화처럼 친숙하게 될 뿐만 아니라, 기계의 성능도 십분 활용할 수 있을 것이다. 구체적으로는 컴퓨터, 로보트, 자동차, 은행의 현금 자동 인출기, 자동 판매기 등 대부분의 기계나 전자제품에 응용가능할 것이다.

3. 영상전화시대의 114안내

현재의 114의 안내는 음성만으로 이루어 지고 있으나, 영상전화시대에는 인물의 등장이 필요할 것이다. 그러나, 전화안내자는 문의전화에 항상 웃는 얼굴로 몇번이고 반복해서 응답하기는 쉬운 일이 아니다. 이때, 합성음성과 합성얼굴영상을 이용한 응답서비스를 사용할 수 있을 것이다.

4. 이산가족찾기 및 미아찾기

남북간의 인적교류가 시작되면, 남북간 이산가족찾기가 실시되리라 생각된다. 그러나, 해방이 된지 50년가까이 되므로 이산가족이 소지하고 있는 사진은 적어도 50년전의 사진이다. 50년전의 모습으로 부터 지금의 모습을 상상하기는 쉬운일이 아니다. 이때, 50년전의 사진으로부터 지금의 모습을 추정·합성하면, 효율적으로 이산가족찾기가 가능하리라 생각된다.

5. 심리학 분야

표정심리학에서는 표정을 이용하여 인간의 행동을 연구하는 경우가 많다. 즉, 여러사람의 동일표정 또는 동일인물의 여러가지 표정, 양자모두가 요구될 때가 있다. 이러한 다양한 요구를 실험대상인물이 전부 만족시키는 것은 어려운 일이다. 이경우 합성표정을 이용하면 이러한 요구는 용이하게 만족시킬 수 있기 때문에, 표정심리학에 있어서는 강력한 실험도구가 될 수 있을 것이다.

6. 범죄수사

범죄수사에 있어서, 현재의 몽타지는 주로 스케치에 의해 정지영상으로 작성되고 있으나, 얼굴합성법을 이용하면 실제사진과 같이 현실감 있는 얼굴영상을 합성할 수 있다. 나아가서, 일단 작성된 몽타지로부터 범인의 특징적인 표정, 옆모습 등을 동영상으로 합성하여, 텔레비전 화면을 통해서 시청자에게 호소한다면 훨씬 효과적인 범죄수사가 가능할 것이다. 또한, 용의자 사진의 정면얼굴과 측면얼굴의 비교 등 동일인물의 식별에도 응용가능할 것이다.

7. 의학분야

성형수술에 있어서 수술후의 모습이 수술전에 상상한 모습과 상이할 때, 의료분쟁을 일으키는 경우가 종종 있다. 이때, 수술후의 모습을 수술전에 미리 합성하므로서 이와같은 시행착오는 줄일 수 있을 것이다. 또한, 수술후의 옆모습, 표정변화 등도 용이하게 볼 수 있을 것이다.

8. 방송분야

어나운서는 방송원고에 따라 뉴스를 전달하게 된다. 이때, 휴먼인터페이스 시스템을 이용하면 방송원고와 얼굴사진만 있으면 텔레비전 화면에 가상의 어나운서가 나타나 뉴스를 전달할 수 있을 것이다. 외국뉴스는 외국인이 우리말로 방송할 수 있으며, 경우에 따라서는 어나운서로서 모나리자의 등장도 꿈만은 아닐 것이다. 또한, 탈랜트의 분장에 있어서도 분장후의 모습을 컴퓨터 시뮬레이션으로 합성해 본후, 가장 적합한 분장을 선택할 수 있을 것이다.

9. 컴퓨터를 이용한 교육(CAE)

컴퓨터를 이용한 교육에서는 피교육자는 숫자나 문자, 도형을 이용하여 컴퓨터 화면에 나타낸 교육내용을 전부 읽지않으면 안된다. 하루종일 피교육자가 컴퓨터 화면을 읽고 있다면, 장시간의 교육은 곤란할 것이다. 이경우, 얼굴합성시스템을 이용하여, 교육내용을 선생님과 같이 대화적으로 설명해준다면 학습효과는 훨씬 증대될 것이다. 나아가서, 역사적 인물이나 원리의 창시자가 직접 설명한다면(아인쉬타인이 나타나서 상대성 원리를 설명한다면) 피교육자의 흥미를 끌 수 있을 뿐만 아니라 교육적 효과 또한 클 것이다. 한편, 청각장애자의 언어교육에 있어서는 입모양의 변화로부터 말을 인식시켜야 하기 때문에, 편

리한 교육기자재가 될 수 있을 것이다.

10. 무인안내시스템

각종건물이나 공공장소에 안내자없이 텔레비전 화면속의 인물이 안내한다면 인건비 절감이 가능할 것이다. 국립민속 박물관에서 실시한 한국인의 얼굴특별전에서는 단군이 관람자에게 초대의 인사말을 하는 안내역을 합성 동영상을 통해서 시도했다.

11. 영화, 광고, 예술분야

현재의 영화는 현존하는 배우만을 이용 제작되고 있지만, 얼굴표정합성법을 신체동작에 까지 확장한다면, 시간과 공간을 초월하여 과거의 배우 또는 외국의 배우와도 자유롭게 연기할 수 있을 것이다. 나아가서, 외국의 배우라고 할지라도 우리말과 일치된 입모양으로 연기하게 될 것이다. 또, 특정인물의 전기를 그리는 영화라면 본인을 직접 등장시키는 것도 먼 장래의 일은 아닐 것이다. 한편, 광고분야에서는 광고의 내용과 관련이 깊은 과거의 인물을 등장시키는 것도 효과가 클 것이다.

Ⅶ. 결 론

얼굴영상을 이용해서 인공현실감을 추구하는 연구에 대해서 소개했다. 이 연구는 얼굴에 나타난 여러가지 정보를 얼굴영상을 통해서 현실감있게 시각화하고, 얼굴영상으로부터 얼굴의 정보를 분석·추출 함으로써 인공현실감을 실현 하려는 방법이라 생각된다. 먼저, 표정정보에 주목하여, 두부의 3차원 동작과 표정정보를 가시화하여 얼굴동영상으로 합성하는 방법에 대해서 살펴보고, 반대로 얼굴동영상으로 부터 이들 정보를 추출·분석하는 방법을 소개했다. 나아가서, 표정정보 이외의 얼굴의 특징정보에 주목하여, 얼굴에 대한 가설을 세우고, 그 가설 하에서, 우리나라 남·여 20세의 평균얼굴, 각지방의 평균얼굴, 일본, 중국, 태국 등의 평균얼굴을 가시화 했다. 또한, 나이정보에 주목하여 연령변화 성분을 추출하여, 선형적으로 가감함으로써 한장의 사진으로부터 일생의 얼굴변화 모습도 시각화 할 수 있음을 알았다. 본고에서 제시된 여러가지 합성영상이 고품질로 다방면에 응용이 가능할 것으로 생각된다. 이외에도 얼굴에 대한 여러가지 정보를 시각화 하여 합성하고, 분석·추출하는 방법을 개발함으로써 고도의 인공현실감을 추구할 수 있으리라 생각된다.

參 考 文 獻

[1] 이남식외 3명 "VR 기술동향 및 산업정책에 관한 연구", 과학기술처 연구보고서 (KRISS-93-078-IR), 1993

[2] 黃瀨通孝, " Virtual Reality 應用戰略", 日本 Ohm사, 1992.

[3] P.Ekman and W.V.Friesen, "Unmasking the Face", Prentics-Hall, 1975.

[4] P.Ekman and W.V.Friesen, "Facial Action Coding System", Consulting Psychologist Press, 1977.

[5] V.Bruce, "Recognising Faces", Lawrence Erlbaum Associates, 1988.

[6] K. Aizawa, H.Harashima and T.Satio, "Model-Based Analysis Synthesis Image Coding System for a Person's Face", Image Comm., vol.1, no.2, pp.139-152,1989.

[7] H.Harashima, K. Aizawa and T.Satio, "Model-Based Analysis Synthesis Coding Videotelephone Images", Trans. IEICE Japan, vol.E72, no.5, pp.452-459,1989.

[8] 相澤 清晴,原島 博, 齊藤隆弘, "構造モデルを用いた畵像の分析合成符號化方式", 日本電子情報 通信學會論文誌, vol.J71-B, no.12, pp. 200-207, 1989.

[9] K.Waters, "A Muscle Model for Animating Three-Dimensional Facial Expression", Comput. Graph., vol.15, no.3, pp.17-24, 1987.

[10] 崔昌石, 原島 博, 武部 幹, "顔の3次元モデルに基づく表情の記述と合成", 日本電子情報通信學會 論文誌, vol.J73-A, no.7, pp.1270-1280,1990.

[11] C.S. Choi, K. Aizawa, H. Harashima

and T. Takebe, "Analysis and Synthesis of Facial Image Sequences in Model-Based Image Coding", IEEE Trans. Circuit System Video Tech., vol.4, no.3, pp.257-275, 1994.

[12] H. Li, P. Roivainen and R. Farchheimer, "3-D motion estimation in model-based image coding", IEEE Trans. Patt. Anal. Machine Intell., vol. 15, no. 6, pp.545-555, June 1993.

[13] 이용동, 최창석, 최갑석 "휴먼인터페이스를 위한 한글음절의 입모양 합성", 통신학회 논문지, vol.94, no.4, pp.614-623, 1994

[14] A. N. Netravali and J. Salz, "Algorithms for Estimation of Three-Dimensional Motion", AT&T Techn. J., Vol. 64, no.2, pp. 335-346, 1985.

[15] B. Horn, "Robot Vision", Chap. 12, pp. 279-298, MIT Press, 1986.

[16] C.S.Choi, T.Okazaki, H.Harashima and T. Takebe, "A System for Analyzing and Synthesizing of Facial Images", IEEE Int. Symp. Circuits & Systems(ISCAS), pp.2885-2888, Singapore, 1991.

[17] 崔 昌石, 岡崎 透, 原島 博, 武部 幹, "主成分分析による顔畵像の基底生成と記述", 日本情報處 理學會 グラフィックスとCAD, 46-7, pp.43-50, 1990.

[18] 박철하, 최창석, 최갑석, "나이변화를 위한 얼굴영상의 주성분분석", 신호처리합동 학술대회 논문집, pp. 703-708, 1992

[19] 박철하, 이원석, 최창석, 최갑석, "얼굴영상을 이용한 성장기 나이변화에 대한 특징성분의 추출과 합성", 통신학회 하계 학술발표회 논문집, pp. 397-400, 1993

筆 者 紹 介

崔 昌 石

1954年 7月 15日生

1978年 2月 홍익대학교 전자공학과 졸업(공학사)

1988年 3月 일본 가나자와(金澤) 대학원 전자정보공학과(공학석사)

1991年 3月 일본 가나자와(金澤) 대학원 전기정보공학과(공학박사)

1988年 4月 ~ 1991年 3月 동경대학 연구원

1984年 1月 ~ 1992年 2月 산업기술정보원(구, KIET) 전자전기실장

1992年 3月 ~ 현재 명지대학교 정보통신공학과 조교수

1992年 일본 전자정보통신학회 논문상수상

주관심 분야 : 영상부호화, 영상처리, 컴퓨터비젼, 컴퓨터 그래픽스