

얼굴 영상의 합성에 관한 연구

박호식* · 배철수*

A Study On Holistic Synthesis Human Face Images

Ho-Sik Park* · Cheol-Soo Bae*

요약

본 논문에서는 전체적으로 묘사된 얼굴의 특징을 이용하여 얼굴 영상을 합성하는 방법을 제안한다. 제안된 방법은 얼굴의 형태를 대조하는 방법을 이용하여 만든 원영상의 작은 부분 영상들을 가지고 얼굴을 재현하였으며, 얼굴 영상에서 2차원적 형태와 구조를 분리함으로서 음영부분과 경계부분을 표현할 수 있었으며, 영상과 영상의 특성을 나타내는 얼굴 부분영역을 조화롭게 재배치함으로서 합성을 가능케 하였다. 실험결과 본 논문에서 제안된 방법으로 합성한 얼굴 영상은 항상 자연스러운 얼굴 영상을 나타냄으로서 제안된 방법의 실효성을 입증하였다.

ABSTRACT

This paper presents a method to automatically synthesized human face images from holistic descriptions. We compactly represent the face set by a small set of prototypes, which can be used in simple ways generated controlled morphings. This becomes possible because separation of 2D-shape and texture provides a faithful, closed and convex representation of images, and smooths the mapping between images and their properties. With this approach, the user watches an images being continuously morphed according to his indications, and the synthesized images always obey the natural physiognomic constraints.

키워드

Facial Recognition, Synthesis, Shape & Texture, Representation

I. 서 론

얼굴합성에 있어서 경찰 조사와 같은 특수한 경우에는 흔히 목격자의 진술에 의존하여 얼굴 영상을 합성해야 할 필요가 있다. 그러나 기존의 얼굴 합성 시스템은 전문가라 할 지라도 만족 할 만한 결과를 얻기가 어려웠다. 그래서 본 논문에서는 컴퓨터에 대한 약간의 지식을 갖춘 초심자의 경우에도 전형적인 표정 혹은 체형 같은 전반적인 특징과 연속적으로 관찰되어진 부분적인 형태를 이용하여 얼굴을 표현하도록 하는 방법을 사용하여 얼굴을 합성할 수 있

는 시스템을 구현하여 자연스러운 얼굴 합성을 이루고자 한다.

본 논문에서 구현하고자 하는 시스템은 영상의 형태와 구조를 분리하여 얼굴의 특성 변화에 반응하는 표본을 정의하였다. 또한, 얼굴의 부분영역특성의 변화에 근거하여 음영부분과 경계부분을 조화롭게 재배치함으로서 얼굴 영상을 합성할 수 있는 방법을 제안한다.

이러한 합성적 표본은 Ullman 과 Basri[8]에 의해 처음으로 이진화 영상에 사용되어졌고, Poggio와 Brunelli^[6]는 형태와 구조를 이용하여 희색등급의 영

* 관동대학교

접수일자 : 2002. 6. 18

상까지 확장하였다. 그 이후로 형태와 구조의 분리는 영상 인식^[2]과 합성에 광범위하게 널리 사용되기 시작했다.^[7,9] 최근에는 Cootes^[5] 등 여러 사람들이 고품질의 표준 모델에서의 형태와 구조의 분리하는 방법을 사용해 오고 있다.

II. 표본

비연속적인 얼굴 부분의 영상 S_{image} 는 $I_{[n \times n]}$ 은 화소 (x, y) 에 대응하는 좌표 $i_{x+n \times (y-1)}$ 의 n 제곱 차원의 벡터 i 로 해석되어질 수 있고, S_{image} 의 부분집합으로 구성된 전체 얼굴 영상(C_{faces})은 여러 가지 변환 요소(자세, 전환, 크기, 조명, 얼굴 표정, 인물식별 등) 내포하고 있다. 전체얼굴(C_{faces}) 안에 포함되는 임의의 S_{image} 의 연속된 지점은 연결선으로 표현할 수 있는 부분과 시각적으로는 동종의 영상을 형성한다고 할 수 있다. 이것은 전체얼굴(C_{faces})은 연관성이 있음을 의미한다.

C_{faces} 의 연관성은 부분 영역에 대응하는 연속적인 위치 변화를 통해 원활한 영상 합성을 가능하게 하였다. 그리고 S_{image} 같은 부분 영역을 이용한 합성은 형태와 특성 변화에 대한 일반적인 해결 방법으로 사용되었고, 본 논문의 합성모델은 형태와 특성 변화뿐만 아니라 얼굴식별 인자의 변화까지 감안한 합성을 가능하게 하여 제어된 추정 변수를 사용하였다.

연속된 변수에서 영상의 형태조성은 반드시 표본을 이용하여야 하고, C_{faces} 는 많은 양의 표본을 필요로 하므로 자료의 압축을 필요로 한다. 또한 경계면 화소의 강도로 인하여 C_{faces} 는 음영의 표시가 제한되게 된다. 일반화된 표본 형태가 없다는 문제로 인하여 합리적인 제어 및 통제를 위하여 묘사된 영상은 반드시 원활한 형태조성을 나타낼 수 있어야 한다. 즉, 묘사된 영상은 반드시 신빙성, 간략함, 정밀함에 부합하여야 한다.

III. 형태와 구조의 분리 표현

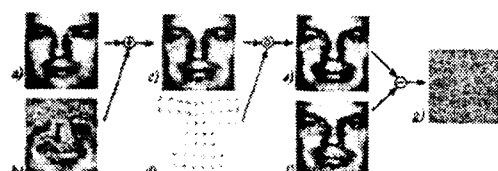
데이터의 감소를 위해서는 압축이 필요로 하며 모든 영상들과의 상관관계 또한 요구된다. 영상간의 연계성을 증가시키는 효과적인 방법으로는 기하학적 변형을 통한 특정 조정이 있다. 일반적인 데이터의 감소를 위하여 저대역 필터 대신에 달리 기하학적 변형 방식을 사용하여 저장과 변형이 용이하게 되었으며, 영상 표현에서 영상을 변형시켰더라도 쉽게 원래의 영상으로 복구시킬 수가 있다.

이것은 표본이 2차원 형태(화소에 따라 조절되는 기하학적 변형)와 구조(기하학적으로 재구성된 영상) 두 부분으로 구성되었다는 것을 알려준다. 영상의 형태와 구조는 분리는 데이터의 감소를 이를 수 있고 PCA(Principal Component Analysis)방식의 압축을 이용한다.

III-1. 표현 연산자

같은 영상 등급의 $n \times n$ 화소의 영상 I_k , ($k = 1, \dots, m$)를 표현하려고 한다. 각각의 I_k 영상은 형태 벡터 $S_k = (\Delta x, \Delta y)_{[n \times n]}$ (R 의 영향으로 I_k 의 모든 화소가 대응됨), 구조 행렬 T_k (I_k 와 R 간의 반응 화소의 강도 차이)를 통해서 등급화된 원 영상 R 과 상관관계를 갖는다. 구조는 보통 형상화된 I_k 에서 R 을 제하면 얻을 수 있다.

$$T_k(x, y) = I_k[x - \Delta x(x, y), y - \Delta y(x, y)] - R(x, y) \quad (1)$$



(a) 원영상 (b) 구조 (c) 구조+원영상 (d) 형태
(e) 재구성 (f) 실제 영상 (g) 재구성 오류
그림 1. 형태와 구조의 재구성

(a) Reference (b) Texture
(c) Texture+Reference (d) Reconstruction
(f) True Image (g) Reconstruction error

Fig 1. Reconstruction from shape and texture

두 연산자 $S_k = shape(I_k, R)$ 과 $T_k = texture(I_k, R)$ 은 형태와 구조 추출을 수행하도록 정의한다. 재구성은 수식의 역변환을 통해 재구성된다.

$$\begin{aligned} rec(R, T_k, S_k) &= T_k[x + \Delta x_{(x,y)}, y + \Delta y_{(x,y)}] \\ &\quad + R[x + \Delta x_{(x,y)}, y + \Delta y_{(x,y)}] \end{aligned} \quad (2)$$

그림 1은 이 과정을 설명해 준다.

III-2. 형태 추정

형태 연산자에게 권한을 주는 두 가지 방법을 연구해 보았다. 첫 번째 방법은 f 를 사용하여 작은 세트의 영상에서 특성을 수동으로 확인하는 방식이다. 변수 $C_k(f \times 2\text{의 크기})$ 는 다음과 같이 사용된다.

1. 모든 특징의 평균값 $\bar{c}_{k,f \times 2}$ 를 찾아낸다.
2. 모든 영상을 연산한다. ($k = 1, \dots, m$)
 - i) 기준에 준하여 각 특성에서 변환시킬 것을 찾는다: $d_k = c_k - \bar{c}$.
 - ii) 모든 화소를 삽입하여 $S_k(x, y)$ 를 만든다.
3. 모든 형태가 일반화된 영상의 평균을 통하여 원영상을 만든다.

$$R = \frac{1}{l} \sum_{k=1}^l [rec^{-1}(I_k, 0, S_k)] \quad (3)$$

$$\begin{aligned} rec^{-1}(I_k, T_k, S_k) &= \\ &rec[I_k, 0, -rec(S_k, 0, S_k)] - T_k \end{aligned} \quad (4)$$

4. 모든 일반화된 영상으로부터 원영상을 제거함으로써 구조를 구한다.

$$T_k = rec^{-1}(I_k, 0, S_k) - R$$

이때 m 영상의 모든 부분을 위한 자동화 과정을 통하여 원 영상 R 을 다루어야 한다. 이 방식은 최적화된 블럭 검색 알고리즘이다. $S_k(x, y)$ 의 변화는 영상 I_k 의 $d \times d$ 와 원 영상 R 사이에 보통화 상호 상호관계를 극대화한다.

형태와 구조 표현이 독특하지 않음으로(형태변화에서의 차이점은 구조 변조를 통하여 상쇄될 수 있다), 다중 분석을 통하여 원활하고 신속한 검색을 사용한다.

반응 화소간의 심도의 차이에서 오는 오류를 제거하기 위해서 구조와 형태를 개량한다. 구조를

$T_k = 0$ 으로부터 시작하여 아래와 같이 진행한다.

1. 형태 $S_k = shape(I_k, R + T_k)$ 을 모든 영상 ($k = 1, \dots, m$)를 위해 연산한다. 복합 단위내 최적합 검색 방식 활용
2. SVD 형태와 구조를 연산한다. 만약 s_k 나 t_k 가 S_k , T_k 의 열 벡터 형식이라면 인수 분해 한다.
- $[s_1] \cdots [s_m] = u_s \sum v_s^T$
- $[t_1] \cdots [t_m] = u_t \sum v_t^T$
3. 축소된 기본배경 $u_s^{(l)}$ 위에 형태를 투사한다. (첫 번째 l 요소)

$$S_k^{(l)} = u_s(S_k^T u_s^{(l)})^T$$
4. 각각의 S_k 와 연관된 구조를 연산한다. 만약 S_k 가 행렬형이라면 $T_k = rec^{-1}(I, O, S_k^{(l)}) - R$ 이다.
5. 축소된 기본배경 $u_t^{(l)}$ 위에 연산된 구조를 투사한다. (첫 번째 l 요소) $t_k^{(l)} = u_t^{(l)}(t_k^T u_t^{(l)})^T$ 매트릭스 형식의 $T_k^{(l)}$ 공식도 염두에 두도록 한다.
6. $S_k = S_k^{(l)}$ 과 $T_k = T_k^{(l)}$ 을 업데이트하고 다시 1단계로 돌아간다.

Beymer^[1]는 감지할 수 없는 데이터 잡음에 대한 해결책으로 이런 정연 된 수렴 순서를 보여준다.

IV. 합성

얼굴 영상을 합성해 내는 것은 높은 등급의 변수를 이용하여 얼굴 영상 공간 C_{faces} (형태와 구조의 표현에 시)에서 위치 변화를 이루어내는 것과 같은 일이라고 할 수 있다. 예를 들어, 코를 크게 만들고 싶거나 화난 표정을 만들고 싶다면 원하는 원활한 위치 변화를 통한 최종 실험 표본을 통해 손쉽게 구할 수 있다.

IV-1. 합성표본

표본은 어떤 규정된 변수의 조합을 나타내는 C_{faces} 의 한 포인트라고 정의할 수 있다. 최종 표본은 다른 평균치와 통제된 최종 표본에 가까운 값을 지닌 몇 개의 변수를 지니고 있다. 최종 표본은 평균적인 변수와 특성이 부과된 경계면의 변수의 요소와 연관되어 있다.

형태와 구조의 분리 표현은 굴곡형의 얼굴형태와 단조로운 할당 작업을 제공한다. 그러므로 변수 I 에 대한 최종 값 n 으로 미리 분류된 표본의 형태와 구조 행렬의 최종 표본 평균치는 :

$$\left\{ \begin{array}{l} S_p(l) = \frac{\sum_{k=1}^n S_k}{n} \\ T_p(l) = \frac{\sum_{k=1}^n T_k}{n} \end{array} \right. \quad (5)$$

큰 n 을 사용함으로써 치우침 없는 결과를 얻을 수 있었다.

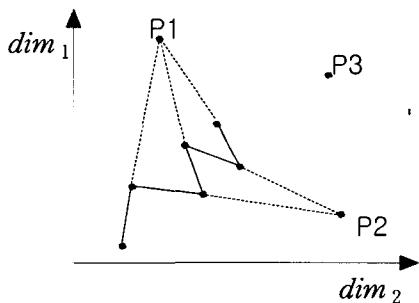


그림 2. p_1 과 p_2 를 이용한 p_3 의 위치
Fig 2. Approximation of p_3 by successive displacements to p_1 and p_2

IV-2. 합성표본의 이용

임의의 얼굴 영상으로부터 시작하여 다음 방법 중 한 가지를 통하여 합성은 이루어진다.

- 최근 제시된 형태와 구조로부터의 지역적인 전개로 최종적 표본 p_1 에 대해 얻어진 x 는 원하는 가치를 지닌다. : $x' = x + \alpha \cdot (p_{(t)} - x)$
변수 1보다 작은 α 는 매번 전시되는 x 와 p_1 의 거리의 일부이다. 그림 2는 대략적인 p_3 에

대한 이 방식을 나타낸다. 몇 가지의 기본적인 표본을 기억하는 것이 매우 효과적이나 합성 위치 변화는 기대치 p_3 에 도달하지는 않는다. 이 방식은 기본 표본의 외피 굴곡을 추정할 수 없다.

- $x' = x + \alpha \cdot p_{(t)}$ 을 통해서 추정은 얻어질 수 있다. 이 방식은 평행 변형법이라고 Beymer 와 Poggio^[2]에 의해 이름지어졌다. 세 번째 영상 $I_3 = rec(R, T_3, S_3)$ 을 변조하기 위하여 $I_1 = rec(R, T_1, S_1)$ 과

$I_2 = rec(R, T_2, S_2)$ 두 샘플간의 변환을 시도하기 위해 사용된다.

- 수식 6을 이용하여 형태 ΔS 와 구조 ΔT 의 차이를 I_1 과 I_2 로부터 구한다.

$$\left\{ \begin{array}{l} \Delta S = shape(I_2, I_1) \\ \Delta T = texture(I_2, I_1) \end{array} \right. \quad (6)$$

- 대략적인 기하학적 일반화를 통하여 I_2 를 변조한다. : $I_4 = rec(R, 0, S_4)$ 와 수식 7

$$\left\{ \begin{array}{l} R' = rec[R, T_3 + rec^{-1}(\Delta T, 0, S_1), S_3] \\ S_4 = rec[rec^{-1}(\Delta S, 0, S_1), 0, S_3] \end{array} \right. \quad (7)$$

- 혼합방식. 1, 2에서 얻어진 방향의 평균치를 이용한다. $x' = (l - \alpha/2) \cdot x + \alpha \cdot p_k$

이 지역전개 중의 일부는 다음 가정중의 ii) 단계와 상관한다.

- 목격자에게 확신하는 자료를 물어보고 첫 번째 대략적 진행에 필요한 기본 외 요소들을 산정 한다
- 이전의 부분적 전개중의 한가지를 이용하여 얼굴 내에서의 변화나 이동에 대해 묻는다.
- 만일 부분적 전개가 성공치 못했을 경우 새로운 기본표본을 선택하여 ii)단계로 돌아간다.

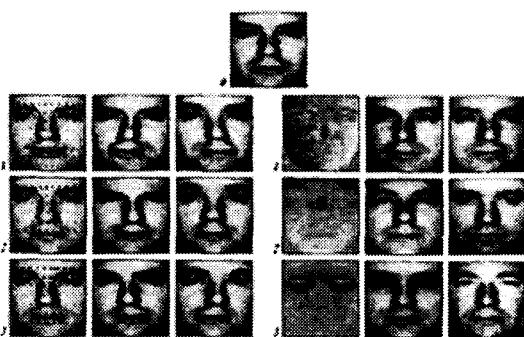


그림 3. 형태와 구조의 3개의 요소와 평균 얼굴 0의 효과
Fig 3. Leading 3 principle component of shpae and texture sets, and their effect over the average face 0

주관적인 기술을 피하기 위해서는 데이터의 선택에 있어서 많은 주의를 요하고 각 변수마다 3개 등급의 중점을 적용하였고 마지막으로 합성 영상은 머리카락이나 안경 등과 같은 추가 내용이나 얼굴표정의 변화 등을 표현하기 위해 공정 후 단계 작업에서 사용된다.

V. 실험결과 및 고찰

1500개의 남자의 디지털화 된 영상(128×128 화소, 32 회색 단계)을 수집하였다. 형태와 구조의 분리는 III-2. 절에서 기술한 바와 같이 적용되었다. 초기 기본 모델의 특성은 기본 방식을 63개의 수동적으로 표기된 특성의 50개의 영상에 적용함으로써 얻어졌다. 상호 형태평가 작용을 통하여 얻어진 PCA 요소는 재구성의 시각적 요소를 배려하여 선택되었다. 10% 미만의 화소가 3 단계보다 더 큰 오류를 나타냈으며 각 공간에 40개의 근본적인 구성요소를 이끌어내었다. 그림 3은 몇몇 전산화된 요소들을 보여준다. 자세와 조명의 효과 및 강세를 측정하고, 영상의 일반화에 불충분한 요소에 대하여 논하였다.

표현에 사용된 요소들은 실험적으로 검증되었고 표현의 충실성은 매우 높이고 집중적으로 실시되었다. 그림 4는 한 특별한 경우를 보여주며 재구성 오류가 낮게 나타남을 보여준다. 표현에서 사용 된 C_{faces} 의 근접성을 확인하기 위해서 이 공간에 임의

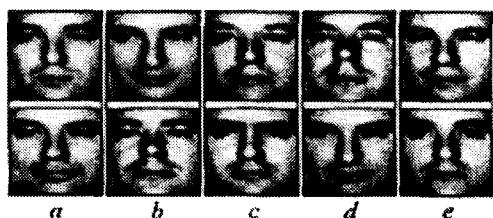
로 지점을 형성시켜보았다. 그림 5가 그 예이다. 이 경우에는 검증된 얼굴 영상이 생성되었으며 C_{faces} 의 표현에 가깝다.



그림 4. 표본의 재구성(좌), 형태와 구조를 이용(중앙) 40 PCA요소의 이용(우)
Fig 4. Reconstruction of an example 9(left) from memorized shape and texture(center) and using 40 PCA components(right)



그림 5. 임의의 표본 일부. 표현들이 비슷함을 나타냄
Fig 5. Some Random examples. Their validity shows that the representation is closed



(a) 입의 너비 (b) 코에서 입의 거리
(c) 눈썹의 분리 (d) 코의 길이 (e) 눈 뜰
그림 6. 특성이 배제된 기본 표본의 일부
(a) Mouth width (b) Nose to mouse distance
(c) Eyebrow separation (d) Nose Length (e) Eye Opening
Fig 6. Some Pairs of elementary prototypes, with properties

C_{faces} 와 그 표본들은 여러 영상들의 평균치이며 그 영상 자체들 역시 여러 영상들의 평균값이다. 그림 6은 표본의 종류를 보여주고 있다. 몇몇 얼굴이 평균 영상을 이루고 있고, 그 자체가 유효한 영상이다.

영상과 선별공간에 대한 할당은 단조롭다는 결론에 도달했다. 왜냐하면 어느 두 지점의 극단적인 평균치라 하더라도 평균치 얼굴에 대한 것이기 때문이다. 그림 6(a)에서 영상은 그림 3에서의 영상 0이다.

오프라인 과정에서는 모든 영상은 자세, 연령, 생김새, 조명과 피부재질에 따라 잘 선별 분리되었고 넓이, 길이, 각도 또 상관적인 얼굴 특징에 대한 측정은 반자동적으로 실시되었으며 기술된 요소에 대한 목록을 완성하였다.

그림 6은 상호 반대되는 최종 기본 표본으로부터 획득된 영상들이다. 그림 7은 표본을 검색하기 위한 두 다른 방법을 이용한 시스템 과정에서의 출력을 보여준다. ○ 과정은 좌측의 같은 얼굴에서 시작되었고 마지막 칸이 원하는 영상이다. 같은 표본의 저장만을 허락함으로써 언제나 실시간으로 시스템을 작동시키고 소량의 메모리를 사용한다.

순수한 평행 형태변형을 이용은 변동 범위를 통제하기 힘들어 실패하였다. 복합방식은 기초 표본의 돌출 부분 및 곡선부분을 삽입하여 좀 더 정확하고 빠른 합성을 가능하게 해주었다.

위의 방식 모두 최종 영상은 항상 자연스러운 인상학적인 형태를 이루게 될 것이다. 사용자가 부분적 변화를 지시했을 때 상관관계로 연결된 영역들이 변화한다. 예를 들면 입의 길이를 증가시켰을 때 표정은 웃음으로 변한다.

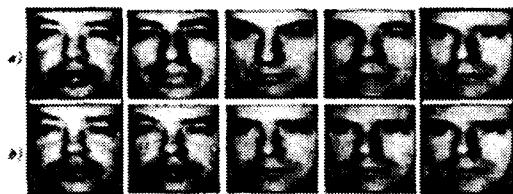
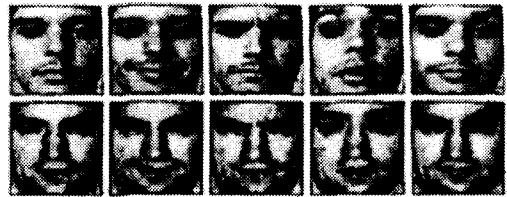


그림 7. 진전된 영상 예
 (a) Using prototype directions
 (b) Using the mixed method
 Fig 7. Examples of controlled image evolution

마지막으로 그림 8은 얼굴 표정을 변화하기 위한 평행 형태변형의 적용의 예이며 이 것은 4.2 절에서 언급한 바와 같이 사후 조정단계로서 사용할 수 있었다.



(위) 원영상 (아래) 외부 삽입에 의한 변형

그림 8. 합성된 얼굴 표정

Fig 8. Synthesized facial expressions. The Top row contains examples of a reference face. In Bottom row, all but lefrmost images are synthesized by application of the exemplified changed

VI. 결 론

본 논문에서는 정면에서 전체적으로 묘사된 얼굴에서 자동적으로 얼굴을 합성하는 방법을 제안하였다. 광범위한 데이터로 인하여 데이터의 축소를 위하여 분리된 형태와 구조의 사용이 필요하였고, 기본 표본의 개념을 단일 특성부분과 남은 다른 영역 부분에 대한 평균치를 이용하여 최종 영상으로 정의하였다. 얼굴의 평균적인 형태와 구조를 이용한 특정 영역의 단조로운 할당으로 기본적인 표본을 구성할 수 있었다.

결론으로, 표본의 충실성과 입체감은 실험을 통하여 검증되었다. 실험은 이 시스템은 합성된 영상이 자연스러운 얼굴 형태에서 어긋나지 않음을 보여줌으로서 제안된 방법의 효율성을 보여주었다.

앞으로의 작업은 표본 영상에 대한 정확한 평균화 작업이 요구되며 특히 자세와 얼굴표정, 명암 및 조명부분에 대한 많은 주의를 필요로 한다. 그러므로 앞으로의 최종적인 표본에 대한 위치의 변화 및 표현에 있어서 좀 더 많은 연구가 필요하다고 본다.

참 고 문 헌

- [1] D. Beymer. "Vectorizing face images by interleaving shape and texture computations." Technical Report AIM 1537, MIT, AI Lab, 1995.

- [2] D. Beymer and T. Poggio. "Face Recognition from one example view." Technical Report AIM 1536, MIT, AI Lab, 1995.
- [3] M. Bichsel and A. Pentland. "Human face recognition and the face images set's topology." *CVGIP: Image Understanding*, 59(2):254-261, March, 1994.
- [4] R. Brunelli and T. Poggio. "Face recognition through geometrical features." In G. Sandini, editor, *Lecture Notes in Computer Science*, pages 792-800, Springer Verlag, 1992.
- [5] T. Cootes, G. Edwards and C. Taylor. "Active appearance models." In *Proc. ECCV'98*, volume II, pages 484-498, 1998.
- [6] T. Poggio and R. Brunelli. "A novel approach to graphics." Technical Report AIM 1354, MIT, AI Lab, 1992.
- [7] D. Rowland and D. Parrett. "Manipulating facial appearance through shape and color." *IEEE Computer Graphics Appl.*, pages 70-76, Sep. 1995.
- [8] S. Ullman and R. Basri. "Recognition by linear combinations of models." *IEEE Trans. PAMI*, 13(10):992-1005, Oct. 1991.
- [9] T. Vetter and T. Poggio. "Linear object classes and image synthesis from a single example image." Technical Report AIM 1531, MIT, AI Lab, 1995.



배철수(Cheol-Soo Bae)

1979년 명지대학교 전자공학과
졸업(공학사)

1981년 명지대학교 대학원 전자
공학과 졸업(공학석사)

1988년 명지대학교 대학원 전자
공학과 졸업(공학박사)

1991년 3월 ~ 현재 산업자원부 공업 기술기반 평가
위원 및 심의위원

1998년 3월 ~ 현재 강원도 정보화위원회 위원

1999년 3월 ~ 2001년 5월 관동대학교 이공대학 학장

2000년 3월 ~ 2002년 2월 동대학교 양양캠퍼스 창
업보육센터 소장

2001년 3월 ~ 현재 해양정보통신학회 강원지부장

2002년 5월 ~ 현재 중소기업 정보화혁신 킨조시업
추진위원

2002.6 ~ 현재 관동대학교 평생교육원 원장

1991년 ~ 현재 관동대학교 정보통신공학부 교수

※ 관심분야 : 디지털신호처리, 영상처리, 신경회로망

저자 소개



박호식(Ho-Sik Park)

1994년 연세대학교 의용전자공학
과 졸업(공학사)

2001년 관동대학교 대학원 전자
통신공학과 졸업(공학석사)

2001년 3월 ~ 현재 관동대학교 대
학원 전자통신공학과 박사과정
재학중

※ 관심분야 : 영상처리, 신호처리시스템, 영상압축