

얼굴의 다양한 포즈 및 표정의 변환에 따른 얼굴 인식을 향상에 관한 연구[☆]

A Study on Improvement of Face Recognition Rate with Transformation of Various Facial Poses and Expressions

최 재 영*
Jae-Young Choi

황보 태 근**
Taeg-Keun Whangbo

김 낙 빈***
Nak-Bin Kim

요 약

다양한 얼굴 포즈 검출 및 인식은 매우 어려운 문제로서, 이는 특징 공간상의 다양한 포즈의 분포가 정면 영상에 비해 매우 흩어져있고 복잡하기 때문이다. 이에 본 논문에서는 기존의 얼굴 인식 방법들이 제한 사항으로 두었던 입력 영상의 다양한 포즈 및 표정에 강인한 얼굴 인식 시스템을 제안하였다. 제안한 방법은 먼저, TSL 모델을 사용하여 얼굴 영역을 검출한 뒤, 얼굴의 구성요소를 통하여 얼굴 포즈를 추정한다. 추정된 얼굴 포즈는 3차원 X-Y-Z축으로 분해되는데, 두 번째 과정에서는 추정된 벡터를 통하여 만들어진 가변 템플릿과 3D CANDIDE 모델을 이용하여 얼굴을 정합한다. 마지막으로, 정합된 얼굴은 분석된 포즈와 표정에 의하여 얼굴 인식에 적합한 정면의 정규화된 얼굴로 변환된다. 실험을 통하여 얼굴 검출 모델의 사용과 포즈 추정 방법의 타당성을 보였으며, 포즈 및 표정 정규화를 통하여 인식이 향상됨을 확인하였다.

Abstract

Various facial pose detection and recognition has been a difficult problem. The problem is due to the fact that the distribution of various poses in a feature space is more dispersed and more complicated than that of frontal faces. This thesis proposes a robust pose-expression-invariant face recognition method in order to overcome insufficiency of the existing face recognition system. First, we apply the TSL color model for detecting facial region and estimate the direction of face using facial features. The estimated pose vector is decomposed into X-Y-Z axes. Second, the input face is mapped by deformable template using this vectors and 3D CANDIDE face model. Final, the mapped face is transformed to frontal face which appropriates for face recognition by the estimated pose vector. Through the experiments, we come to validate the application of face detection model and the method for estimating facial poses. Moreover, the tests show that recognition rate is greatly boosted through the normalization of the poses and expressions.

□ Keyword : Face Recognition, Facial Pose, Facial Expression, 3D Face Model

1. 서 론

출입 통제 시스템 등의 신원확인이 필요한 곳에서는 지문 인식, 홍채 인식, 음성 인식 등의 생

체 인식과 함께 얼굴 인식이 매우 중요한 비중을 차지하고 있다. 이는 다른 접촉 인식 방법들에 비해 얼굴 인식이 사용상의 거부감과 불편함이 적고, 사용자가 알지 못하는 사이에 얼굴을 인식할 수 있는 지능형 무인 감시 시스템, 범죄 용의자 검색 시스템 등의 분야에 있어서는 기존의 인식 방법들의 사용 제한을 보완해 줄 수 있기 때문이다. 이러한 얼굴 인식은 정확도의 성능 향상을 위해 그 동안의 많은 연구가 이루어지고 있고 이로 인하여 응용 시스템으로 적용할 만큼의 높은 인

* 정 회 원 : 경원대학교 소프트웨어연구소 책임연구원
jy1001@kyungwon.ac.kr(제1저자)

** 중신회원 : 경원대학교 소프트웨어대학 부교수
tkwhangbo@kyungwon.ac.kr(공동저자)

*** 중신회원 : 경원대학교 소프트웨어대학 교수
nbkim@kyungwon.ac.kr(공동저자)

☆ 이 논문은 2003년도 대학 IT 연구센터 육성지원사업비의 지원을 받아 연구되었음.

식률을 가지게 되었지만, 대부분이 정면 얼굴에 대한 연구로서 실제세계에서는 조명 및 포즈 변화 등 얼굴의 다양한 변화 때문에 그 인식률이 지문이나 홍채에 비해 낮게 나타나며 현재까지도 이를 보완하는 기술의 요구가 절실하다[1].

또한 얼굴 포즈 및 표정의 변화가 있는 경우 인식의 차이가 크다. 즉, 같은 사람의 영상일지라도 포즈가 다른 두 영상은 공간 영역에서 큰 차이를 나타내기 때문에 이를 입력받는 기존의 정면 영상 인식 시스템은 큰 에러를 발생할 확률이 크다. 이에 본 논문에서는 다양한 포즈 및 표정을 갖는 얼굴 영상이 입력되면 얼굴의 포즈를 기하학적으로 추정하고 기본 표정을 분석한 후에 이를 얼굴 인식에 적합한 정면 영상으로 변환하는 방법을 제안한다.

제안한 시스템은 검출된 얼굴 구성 요소와 얼굴 영역의 기하학적 구조를 이용하여 입력 영상에 대한 포즈를 추정하며, 이를 통하여 얻은 포즈의 방향(direction)을 분해하여 좌우회전(yaw), 상하회전(tilt), 그리고 얼굴 자체 회전(roll)값을 수치로 산출한다.

산출된 수치는 얼굴을 모델링하기 위한 가변 템플릿의 회전에 사용되며, 이를 통하여 정합된 얼굴은 정면 영상으로 합성 복원한다. 이때, 얼굴 표정의 기본 요소를 분석하여 기준치 이상으로 변형된 얼굴 표정도 정규화 시켜준다. 따라서, 기존의 정면 얼굴 인식 시스템뿐만 아니라 다양한 포즈 및 표정을 접해야 하는 로봇 비전 시스템에서의 얼굴 인식 성능 향상 및 모바일 폰에서 다양한 얼굴의 검출 및 인식을 위한 기반 기술에도 기여할 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 제안한 방법인 기하학적 구조를 이용한 얼굴 포즈 추정방법을 제시하며 제 3장에서는 이를 이용한 포즈 및 표정의 정규화 변환 방법에 대하여 설명한다. 제 4장에서는 제시한 방법의 실험 및 결과를 설명하고, 마지막으로 제 5장에서 결론을 맺는다.

2. 기하학적 관계를 이용한 얼굴 검출 및 포즈 추정

현재까지 얼굴 인식을 위한 다양한 방법이 제안되고 있지만 대부분의 연구가 정규화된 정면 영상만을 입력 데이터로 제한함으로써, 실제로 얼굴의 회전이 빈번하게 일어나는 일반 영상에서의 얼굴 인식률은 상대적으로 낮은 결과를 보이고 있다. 동영상 감시 시스템의 경우, 일반적으로 CCD 카메라로부터 획득한 4~5장 정도의 스틸 사진을 확보하여 얼굴을 인식한 후 이중 마할라노비스(Mahalanobis) 거리나 유클리디언(Euclidean) 거리가 가장 적은 얼굴을 인증하는 방법이 쓰이는데, 사전에 포즈의 상태를 추정한다면 4~5장 모두를 얼굴 인식 모듈에 수행할 필요 없이 포즈의 회전각이 정면영상과 가장 가까운 사진만을 선택하여 처리함으로써, 수행시간을 단축시키고 잘못된 입력 영상으로 인한 오인식을 줄일 수 있다.

본 장에서는 TSL 모델을 이용하여 얼굴을 검출하고 검출된 얼굴 영역 내에서 얼굴 구성 요소로 가정할 수 있는 홀들을 찾아내며, 이들의 기하학적 관계를 규명하여 최종 구성 요소인 양 눈과 입을 찾아낸다. 찾아진 구성요소는 포즈를 추정하는데 사용되며, 추정된 포즈를 통하여 좌우 및 상하, 회전각도를 산출하게 된다.

2.1 얼굴 검출

TSL 색상 공간 중 T-S 공간은 조명과 독립적으로 하나의 군집을 이루고 있고 그 분포가 밀집되어 있어 사람의 얼굴색 추출을 위한 방법으로 적합하다.

식 (1)은 TSL 색상 모델의 계산식이며, T는 색조(tint), S는 선명함(saturation), 그리고 L은 휘도(luminance)를 나타낸다[2]. 본 논문에서 T-S 정보는 얼굴 및 입술 검출에 이용하며, L 정보는 조명을 분석하고 보정하는데 사용한다.

$$T = \begin{cases} \arctan(r'/g')/2\pi + 1/4, & g' > 0 \\ \arctan(r'/g')/2\pi + 3/4, & g' < 0 \\ 0, & g' = 0 \end{cases} \quad (1)$$

$$S = [9/5(r'^2 + g'^2)]^{1/2},$$

$$L = 0.299R + 0.587G + 0.114B$$

, where $r' = (r-1/3)$, $g' = (g-1/3)$
 $r = R/(R+G+B)$, and $g = G/(R+G+B)$

식 (1)을 사용하여 한국인의 피부 샘플을 학습 시킨 후에 학습한 얼굴색 공간을 가지고 얼굴 영역을 분류하는 방법은 TSL 색상 공간이 타원형 분포를 이루고 있다는 점에 착안하여, 식 (2)-(3)과 같이 타원형 분포의 거리 측정에 적합한 마할라노비스 거리 측정법을 이용한다.

$$P(X(i,j) | K) = \frac{\exp[-\frac{1}{2}((X(i,j) - m_K)^T C_K^{-1} (X(i,j) - m_K))]}{2\pi |C_K|^{1/2}} \quad (2)$$

$$[\lambda_K(i,j)]^2 = [X(i,j) - m_K]^T C_K^{-1} [X(i,j) - m_K] \quad (3)$$

하지만, T-S 색상 모델을 통하여 검출된 영상 또한 얼굴 내부의 얼굴색뿐만 아니라 주위의 색상이 얼굴색과 유사하여 원하지 않는 부분이 검출될 경우가 있다. 이런 경우 레이블링을 시행하여 가장 큰 영역을 선택함으로써 최종 얼굴 영역을 찾는다. 일반적으로 잡음 제거를 위한 방법으로 모폴로지(morphology)의 열림(opening) 연산 등의 방법이 있으나, 모폴로지 방법은 본 논문의 다음단계에서 시행할 홀 검출과정에서 홀의 변형을 가져올 수 있으므로 그림 1과 같이 레이블링 방법을 수행하였다.



(a) 원영상, (b) 얼굴 영역 검출 결과, (c) 레이블링 결과
 (그림 1) 레이블링을 통한 얼굴 영역 선택

2.2 각도별 광원 분석을 통한 조명 보정

색상 모델을 이용하여 얼굴 영역을 검출할 때 발생하는 가장 큰 문제점은 조명에 의해 잘못된 결과를 유발하는 것이다. 즉, 영상에 따라 일정하지 않은 조명의 밝기는 결과적으로 다른 얼굴색을 가진 화소값으로 나타나며, 이는 다른 영역으로 분류될 수 있기 때문에 조명 보정을 통하여 조명에 대한 영향을 최소화하도록 한다. 특히, 얼굴 전체의 무게중심을 검출된 얼굴색의 화소분포로서 산출하는 본 논문에 있어서 확장된 얼굴 영역 검출은 필수적이다. 이에 본 논문에서는 얼굴의 각도별 조명의 영향을 분석하여 광원을 찾아낸 다음 조명의 영향을 받은 부위에는 조명 보정을 통하여 얼굴색을 검출하는 방법을 제안한다.

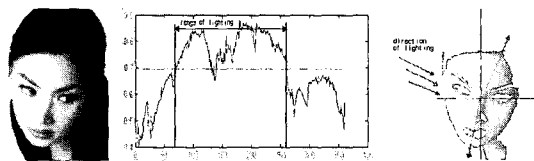
$$L(\theta) = \frac{\sum f(x,y)}{N_\theta}, \text{ where } f(x,y) \in L_{face} \quad (4)$$

$$\theta = \arctan[(f_y - CL_y)/(f_x - CL_x)] \times 180/\pi \quad (5)$$

CL : 얼굴색 영역 L_{face} 의 무게 중심

N_θ : 각도 θ 에 해당하는 L_{face} 의 화소수

해당 각도의 화소 밝기값은 TSL 색상 공간의 계산식을 나타낸 식 (1)의 휘도값 L을 사용하였으며, 식 (4)-(5)와 같이 얼굴 영역의 중심을 기준으로 전체 각도별 조명 평균값을 산출한 후에 이를 분석하여 광원의 각도를 찾아낸다.



(a) 입력 영상 (b) 조명 분석 (c) 조명 영향 영역
 (그림 2) 조명 분석 결과

그림 2는 입력 영상을 분석하여 조명의 각도를 찾아낸 결과를 보여준 것으로서, 얼굴 영상의 중심좌표를 기준으로 산출한 각 각도별 밝기값의 평

균들이 전체 밝기의 평균 이상이 되는 각도 범위는 조명의 영향을 받은 부분이 된다. 따라서 산출된 각도 범위에 해당하는 얼굴 영역은 밝기값을 보정하여 얼굴 영역을 재추출함으로써 일반적으로 시행하는 영상전체에 대한 조명 보정이 아닌 그림 3과 같이 필요한 피부 영역만을 보정하게 된다.



(a) 조명 보정 전 (b) 조명 보정 후
(그림 3) 조명 보정 결과

2.3 얼굴 영역 내에서의 홀 검출

얼굴색 검출 방법을 통하여 얻어진 얼굴 영역 내부에는 미검출된 화소가 발견될 수 있으며, 본 논문에서는 이 영역들을 홀(hole)이라 명하였다. 일반적으로 얼굴 영역을 획득한 후에 얼굴 내부의 특징점을 찾는 방법은 기하학적 방법이나 템플릿 방법 모두 에지 검출 방법을 전처리로서 사용한다. 에지 검출이 주로 사용된 이유는 사람의 얼굴을 구성하는 요소인 눈썹, 눈, 코밑, 입 등이 모두 수평성분을 유지하는데 그 이유가 있다. 하지만, 현재 검출하려는 입력영상은 다양한 얼굴 방향을 가지고 있기 때문에 이들의 얼굴 구성요소도 수평성분이라고 볼 수 없다. 따라서, 방향정보를 전혀 가지고 있지 않는 홀은 얼굴색 영역에서 제외된 얼굴내의 요소이므로 이들을 일단 눈, 코, 입 등의 영역이라고 가정하고 이들 서로간의 기하학적 관계를 비유함수로 연산하여 최종 얼굴 요소를 추출한다. 홀을 탐색하는 방법은 그림 4의 (a)와 같이 4방향으로 탐색을 거치게 되는데, 그 중에서 3-방향 이상이 막혀있다면 홀로 판정하게 된다. 예를 들어, 그림 4의 (b)와 같이 주변의 3-방향이 막혀있는 경우 이 영역은 홀로 판단된 경우이고, (c)의 경우는 2-방향만이 막혀 있기 때문

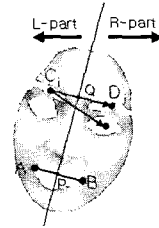


(a) 홀 탐색을 위한 4-방향성 (b) 홀 영역 (c) 비 홀 영역
(그림 4) 홀 탐색 방법

에 홀로 판단할 수 없는 경우이다.

2.4 얼굴 구성 요소 선정

얼굴 구성 요소 후보들이 추출되면 세선화를 통해 얻어진 입의 정보를 이용하여 그림 5와 같이 얼굴의 기하학적 관계를 구성한다. 즉, 입의 양 끝점을 지나는 직선과 수직인 직선을 가상으로 정한 후 이를 중심으로 왼쪽에 있는 구성 후보를 L-part로, 오른쪽에 있는 구성 후보를 R-part로 놓고 서로 간의 관계를 고려하여 최종 구성 요소를 선정한다.



(그림 5) 얼굴 구성 요소의 기하학적 관계

구성 요소의 관계는 L-part에 있는 홀의 중심 C_i , 하나와 R-part에 있는 홀의 중심 D_j , 하나씩의 조합으로 계산되며, 식 (6)의 세가지 조건을 만족하는 (C_i, D_j) 쌍이 최종적으로 선택되는 좌우 눈이다.

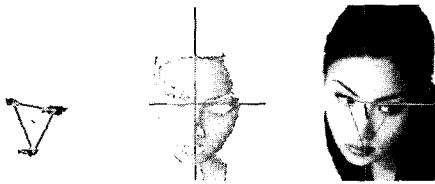
$$[|AB| \approx |C_i D_j|] \text{ and } [\overline{AB} // \overline{C_i D_j}] \text{ and } [|PQ| < \epsilon] \quad (6)$$

식 (6)은 눈을 이루는 쌍과 입의 선분은 유사

한 길이를 가지고 있으며 평형을 이루어야 함을 나타내는데, 일반적으로 이러한 조건을 만족하는 양 눈과 양 눈썹간의 관계에서 입에 가까운 쪽에 있는 양 눈을 선택하기 위하여 입으로부터 일정 거리를 가지는 구성요소를 택한다.

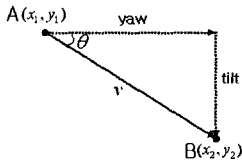
2.5 얼굴의 포즈 추정 벡터

본 연구에서 제안한 얼굴 영역 검출 및 구성 요소 추출 과정에서 얻어진 순수 얼굴 영역의 중심 좌표로부터 얼굴 구조 요소(양 눈, 입)들의 삼각형 무게 중심을 연결하여 벡터를 얻을 수 있으며, 이 벡터를 통하여 얼굴의 포즈 추정 값을 산출한다.



(a) 구성요소 중심 (b) 얼굴 중심 좌표 (c) 포즈 추정 결과
(그림 6) 얼굴의 포즈 추정 벡터

그림 6과 같이 포즈 추정을 통하여 산출된 벡터를 살펴보면, 벡터 방향은 얼굴 포즈의 방향이고, 길이는 포즈 방향으로의 이동치를 의미하며, 그 길이를 그림 7과 같이 분석하여 좌우 회전과 상하 회전에 대한 각 성분별 수치를 각각 산출할 수 있다.



(그림 7) 각도 산출을 위한 벡터 분석

그림 7에서 $A(x_1, y_1)$ 는 얼굴 영역의 중심좌표이고 $B(x_2, y_2)$ 는 얼굴의 구성 요소로 이루어진

삼각형의 무게중심을 나타내며, 포즈 추정 벡터 $v = \overline{AB}$ 를 수평방향의 v_{yaw} 와 수직방향의 v_{tilt} 로 분해하기 위하여 식 (7)에서 식 (10)까지 대입하면 각각 성분에 맞는 결과를 얻을 수 있다.

$$v = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad (7)$$

$$v_{yaw} = \frac{v \cdot \cos \theta}{half_face_width} \cdot 90 \quad (8)$$




$$, where \theta = \arctan\left(\frac{|y_2 - y_1|}{|x_2 - x_1|}\right)$$

$$v_{tilt} = \frac{v \cdot \sin \theta}{half_face_height} \cdot 90 \quad (9)$$

$$\angle roll_{face} = \arctan\left(\frac{|y_{Reve} - y_{Level}|}{|x_{Reye} - x_{Level}|}\right) \times 180/\pi \quad (10)$$

표 1은 포즈값을 산출한 결과로서, 좌우 회전의 L과 R은 각각 왼쪽방향과 오른쪽 방향을 의미하며, 상하 회전의 U와 D는 각각 위 방향과 아래 방향을 의미한다. 얼굴 회전에서의 부호는 시계의 3시 방향을 기준으로 시계 반대방향으로 기울어졌을 때에는 +이며, 시계방향으로 기울어졌을 때에는 -값이다. 이는 실제 측정하는 각도와 같은 원리로 하였다.

(표 1) 포즈 값 산출

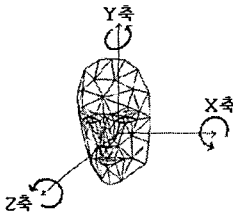
영상 성분				비고
좌우 회전(yaw)	R 18.6°	R 5.8°	R 2.1°	Y축 회전값
상하 회전(tilt)	D 3.1°	D 21.6°	D 4.5°	X축 회전값
얼굴 회전(roll)	-18.4°	-10.8°	+8.9°	Z축 회전값

3. 얼굴 정합 및 포즈 / 표정 변환

얼굴을 모델링 할 경우는 특징점의 개수가 많을 수록 정밀한 얼굴을 모델링할 수 있지만 수행속도

와 연산의 복잡도가 증가하는 반면, 특징점의 개수가 적으면 빠른 연산 속도를 얻을 수 있지만 모델의 질은 그만큼 감소하게 되므로 특징점의 개수는 응용하는 분야 따라 적절히 결정해야 한다.

본 논문에서는 J. Ahlberg가 기존의 CANDIDE를 향상시킨 CANDIDE-3 3D 와이어 프레임 모델을 사용하는데, 이 형상 모델은 Linköping University에서 모델 기반 영상 코딩과 컴퓨터 애니메이션의 목적으로 처음 만들어 졌다[3]. 그림 8은 113개의 점들과 183개의 삼각형의 집합으로 이루어진 CANDIDE-3 얼굴 모델이다.



(그림 8) CANDIDE 모델과 3차원 공간상에서의 회전축

포즈 정규화는 3차원 얼굴 모델을 정합시키기 위한 얼굴 구성요소 외의 몇가지 특징점을 추가로 찾고, 이 값을 이용하여 얼굴 모델과 입력 영상을 정합 시킨다. 정합된 입력 영상은 2장에서 추정된 포즈의 역값을 이용하여 그림 8과 같이 X, Y, Z축의 회전 조합을 통하여 정면 영상이 되도록 형상 모델과 이에 매핑된 텍스처를 변환한다.

3.1 특징점 추출

본 논문에서는 얼굴의 특징점을 찾기 위하여 템플릿 매칭방법을 사용하는데, 이 방법은 기하학적 방법보다 정확도가 뛰어나며, 눈, 코, 그리고 입의 대략적인 위치를 알고 있는 경우 기존의 단점으로 여겨졌던 탐색 속도를 향상시킬 수 있어 널리 쓰이고 있다[4].

본 논문에서 추출하고자 하는 추가적인 특징점은 눈썹의 양끝, 눈의 양끝과 상하, 코의 양끝,

입의 양끝과 상하, 턱, 그리고 얼굴 폭이다.

하지만 템플릿 매칭 단점중의 다른 하나는 찾고자 하는 대상과 미리 설정한 마스크가 매우 유사하여야 하는데, 현재 본 연구에서 사용하는 입력데이터는 다양한 포즈를 가지고 있으므로 이에 해당하는 템플릿 마스크를 사전에 모두 준비하고 있다는 것은 매우 비효율적이며, 템플릿을 적용할 때 적합한 마스크를 선택한다는 것도 연산의 복잡도가 높아지는 현상을 초래한다. 따라서, 본 연구에서는 그림 10의 (a)와 같이 기본 정규화된 템플릿 하나를 가지고 이미 추정된 기하학적 특징(포즈 분석에 의한 회전 값)으로 그림 10의 (c)와 같이 변형시키는 가변 템플릿 방식을 제안하여 템플릿 매칭을 시도하였다.

3.2 추출된 특징점을 이용한 얼굴 정합

3.1절에서 추출된 특징점들과 2장에서 추정된 포즈를 이용하여 얼굴 모델을 회전시켜 입력 영상과 정합시킨다. 이때, 형상 모델을 회전시키는데 있어서 회전운동 $R = R(r_x, r_y, r_z)$ 은 각각 X축, Y축, 그리고 Z축을 중심으로 한 회전으로 분해할 수 있다. 또한, 회전과 이동 운동은 동차 형식(homogeneous form)을 사용하면 편리하므로 이를 위해 3차원 위치 벡터에 1을 더하여 재정의 하였다.

여기서, 얼굴 영상에 대한 정합을 위한 크기 변화인 s 는 X축, Y축, Z축 모두에 대하여 각각 변화의 배율이 달라져야 하므로 크기 변화의 등차 변환 행렬을 사용한다.

그림 10의 (d)와 (e)는 이와 같은 방법을 이용하여 얼굴을 3차원 얼굴 형상 모델에 정합한 결과이다. 기본적인 얼굴 형상 모델의 회전은 3.1절에서 구한 포즈 추정값에 의하여 R 을 회전시켰으며, 얼굴의 크기를 고려하여 s 를 조절하고 추가적으로 추출된 특징점을 통하여 특징점 위치를 세부 조정하였다.

3.3 정면 얼굴로의 포즈 변환

본 절에서는 2장에서 추정된 포즈값을 토대로 얼굴의 다양한 포즈를 정면의 얼굴로 정규화하는 과정을 제안한다. 여기서 사용한 정규화라는 용어는 다양한 각도로 움직인 포즈값 즉, 그림 8의 X축 회전(tilt), Y축 회전(yaw), 그리고 Z축으로 회전(roll)된 포즈값을 각각 0으로 보정함으로써 회전된 얼굴 템플릿을 정면의 얼굴로 합성하여 복원하는 방법을 일컫는다. 얼굴의 템플릿은 형상 모델과 텍스처로 구성되며, 얼굴 템플릿의 변형은 형상 모델의 꼭지점 좌표 값의 변화와 세 개의 꼭지점으로 이루어진 삼각형 텍스처의 재 합성으로 이루어진다. 즉, 삼각형 A_a, B_a, C_a 에서 삼각형 A_b, B_b, C_b 로의 화소값 이동은 목적 영상에서 식 (11)과 같이 목적 영상의 s_i 와 t_i 를 구한 후에, 이를 원영상에 대입하여 가져오거나 하는 화소값 q_a 를 구한다.

$$q_b = s_i(B_b - A_b) + t_i(C_b - A_b) + A_b \quad (11)$$

즉, $A_b = (x_0, y_0)^T$, $B_b = (x_1, y_1)^T$, 그리고 $C_b = (x_2, y_2)^T$ 라 하면, $q_b = (x_i, y_i)^T$ 의 위치를 알려주는 비례계수인 s_i, t_i 는 식 (12), 식 (13)과 같이 구할 수 있고, 이를 식 (14)에 대입하여 q_a 를 구하는 것이다.

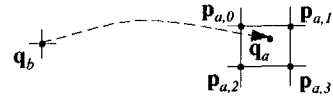
$$\begin{bmatrix} x_i - x_0 \\ y_i - y_0 \end{bmatrix} = \begin{bmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \end{bmatrix} \begin{bmatrix} s_i \\ t_i \end{bmatrix} \quad (12)$$

$$\begin{bmatrix} s_i \\ t_i \end{bmatrix} = \begin{bmatrix} x_1 - x_0 & x_2 - x_0 \\ y_1 - y_0 & y_2 - y_0 \end{bmatrix}^{-1} \begin{bmatrix} x_i - x_0 \\ y_i - y_0 \end{bmatrix} \quad (13)$$

$$q_a = s_i(B_a - A_a) + t_i(C_a - A_a) + A_a \quad (14)$$

이때, 템플릿의 텍스처가 비트맵으로 저장되는 경우 텍스처의 화소의 위치는 항상 정수가 되어야 하지만 이동되어 변형된 템플릿 텍스처의 위치는 꼭 정수가 되라는 보장이 없기 때문에 그림

9와 같이 정수 위치에 대응하는 이동전 텍스처의 인접하는 4개의 화소 값을 이용하여 보간하는 역방향 워핑 방법을 수행함으로써 선형도를 만족시키는 화소 값을 구한다.

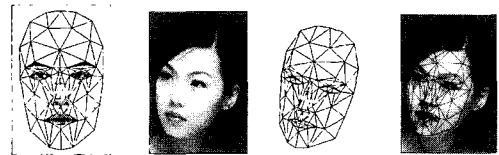


(그림 9) 역방향 워핑을 이용한 화소값 보간

그림 9는 변형하고자 하는 영상 f_b 의 화소값 q_b 를 원 영상 f_a 의 화소 $p_{a,i}$ ($0 \leq i \leq 3$) 사이에 각각 α 와 β 의 비율로 위치한 q_a 를 식 15와 같이 계산하여 화소값을 얻는 역방향 워핑 방식을 나타낸 것이다[5].

$$f_b(q_b) = \sum_{i=0}^3 w_i(\alpha, \beta) f_a(p_{a,i}) \quad (15)$$

그림 10은 이러한 과정을 걸쳐 정규화 방향, 즉 정면 영상으로 재 합성된 결과이다. 이는 2장에서 추정된 포즈 값에 의하여 정면으로 복원된 경우인데, 그림 10의 (f)와 같이 얼굴 전체는 정면으로 변형되었으나 눈동자의 위치가 보정되지 않아 측면을 향하고 있는 것처럼 보인다.



(a)기본템플릿 (b)입력영상 (c)가변템플릿 (d)정합결과



(e)텍스처 매핑 (f)포즈 정규화

(그림 10) 얼굴 정합 및 포즈 정규화

3.4 표정의 분석

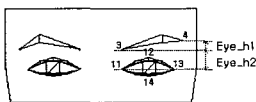
본 논문에서는 얼굴 정합 과정에서 얻은 특징점의 이동 변위를 FACS(Facial Action Coding System)에서 정한 표정 유닛(AU : Action Unit)의 범주 내에서 판단하여[6], 그림 11 및 그림 12와 같이 만약 특징점이 기준으로 설정한 입개 값 이상의 범위로 이동하였을 경우 이를 정상적인 위치로 보정함으로써 개인의 얼굴 특성을 손상시키지 않고 얼굴 인식에 적합한 정규화된 얼굴을 만들도록 하였다.

AU의 움직임에 의한 얼굴 표정 변화는 얼굴 영상에 정합된 형상 모델의 꼭지점의 위치에 그대로 반영되며, 정규화된 기본 형상 모델 템플릿을 기준으로 변화된 꼭지점의 위치를 알아낸 후 그 차이를 비교하여 표정을 알아낸다. 즉, 식 (16)에 사용되었던 형태 유닛 S 의 이동 변위 o 를 측정하여 표정을 분석한다.

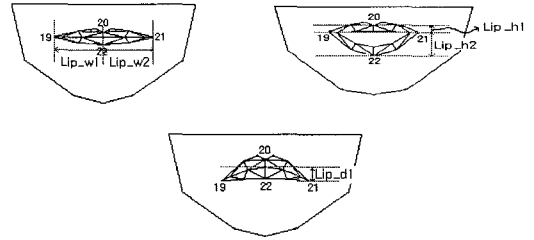
$$g = \bar{g} + oS + aA \quad (16)$$

식 (16)에서 구하고자 하는 값 g 를 위하여 모델의 표준 형태 \bar{g} 에 변형된 형태 유닛 S 와 표정 유닛 A 를 추가하며, o 와 a 는 각각 형태와 움직임을 조절하는 변수로서 0에서 1사이의 값이다.

예를 들어, 그림 11에서 왼쪽 눈썹 안쪽 끝점 F_3 의 이동 변위 $o_{F_3}(x, y, z)$ 는 눈의 양 끝점 F_{11} 과 F_{13} 을 잇는 기준선으로부터의 y 성분 거리 Eye_h2 로서 기본 형상 모델에서 최초 주어졌던 변위와의 차이를 통하여 얻을 수 있다. 특징점 F_4 도 마찬가지로 Eye_h1 과 Eye_h2 의 합산 거리를 기본 형상 모델과의 차이와 비교하여 이동 변위를 추정한다.



(그림 11) 상부(눈썹)의 표정 분석



(a) 입술을 길게 만듦(AU-(23+24)) (b) 입을 벌림(AU-(26,27)) (c) 입 끝을 내림(AU-15)
(그림 12) 하부(입)의 표정 분석

추정된 변위에 의하여 원 위치로의 복원은 실제 해당하는 CANDIDE의 꼭지점들의 그룹이동을 의미하며, 그림 11의 왼쪽 눈썹을 움직이고 할 때에는 aA_{l-eye} 내의 형태 유닛들 x, y 좌표가 이동되는 것이며, 오른쪽 눈썹은 aA_{r-eye} 내의 형태 유닛들이 움직이는 것이다.

하부에 해당하는 입의 표정 분석 및 정규화도 마찬가지로 방법으로 수행된다. 각 특징점의 매핑 변위는 기본 템플릿에 추가된 o 를 통하여 분석되며, 마찬가지로 방법으로 이에 해당 형태 유닛들의 집합인 aA 의 이동으로 원위치로 복원된다.

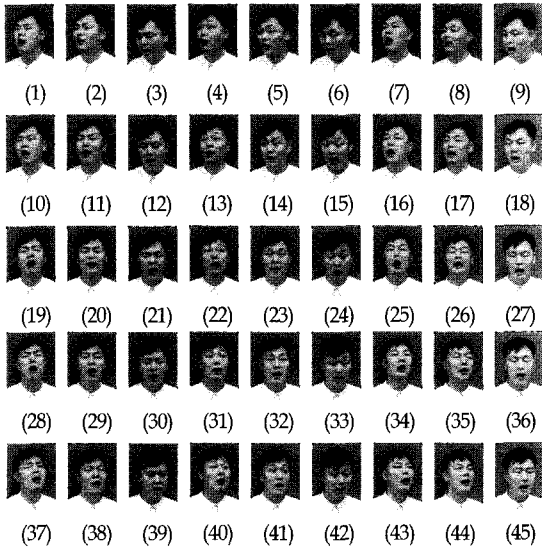
4. 실험 및 결과 분석

그림 13은 다양한 얼굴 각도를 촬영한 포즈와 표정이 혼합되어있는 얼굴 영상을 획득한 것으로서, 실험의 알고리즘의 구현은 도플 및 분석을 위하여 MATLAB 6.0과 Visual C++ 6.0을 사용하였다.

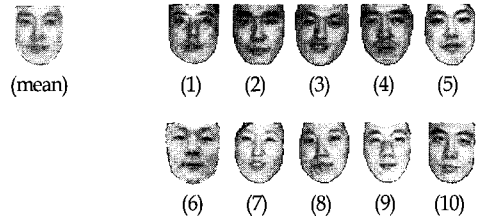
4.1 실험 방법

포즈 및 표정 정규화 실험은 2단계로 나누어서 1단계에서는 얼굴 포즈만을 추정하여 얼굴을 정면 영상으로 변환시키는 실험만을 하며, 2단계에서는 얼굴의 포즈뿐만 아니라 표정까지 변환하는 최종 정규화 영상을 만들어낸다.

또한, 본 논문에서 제시한 방법의 최종 목표인 얼굴 인식에 관한 성능 평가를 위하여 그림 14와



(그림 13) 다양한 포즈 및 표정 영상



(a) 평균 얼굴 (b) 학습 얼굴

(그림 15) 주성분 분석을 위한 얼굴 영상들

$$\Omega^T = [\omega_1 \omega_2 \dots \omega_{M'}] \quad (17)$$

$$\omega_k = u_k^T \left(\Gamma - \frac{1}{M} \sum_{n=1}^M \Gamma_n \right) \quad \text{for } k=1, \dots, M' \quad (M' \ll M)$$

$\Gamma_n (n=1, \dots, M)$: N차원의 벡터로 표현되는 M 개의 벡터 집합

u_k : 각각의 열에 직교한 기저벡터

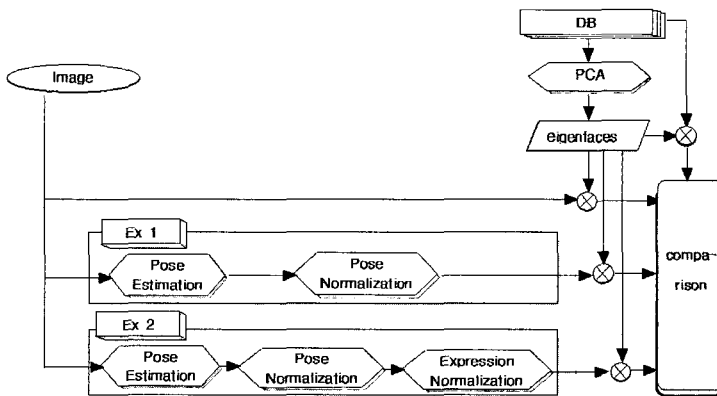
같이 다양한 포즈를 가진 입력 영상을 원래대로 인식했을 경우와 실험 1과 실험 2를 각각 거쳤을 때의 인식률을 비교해 본다. 실험 1은 다양한 포즈를 가진 입력 영상을 포즈값 계산 후에 포즈만 정규화 시킨 방법이며, 실험 2는 실험 1의 과정에 표정 정규화 단계가 더 추가된 것이다. 일반 방법과 두 가지 실험을 통하여 나온 3개의 영상은 주성분 분석 방법을 이용하여 비교되며, 이 때 산출되는 인식률 및 인식을 위한 거리 값의 차이를 그래프로 도식하여 비교 분석한다.

주성분 분석을 위하여 그림 15와 같이 대상 인물을 포함한 10명의 정면 얼굴을 학습시켜 식 (17)과 같은 고유벡터를 산출했으며 이들간의 관계를 통하여 구하여진 고유벡터들의 거리값 차이를 이용하여 식 (18)과 같이 정규화된 얼굴의 오차여부를 측정하였다.

$$\varepsilon_k = \|y_i - y_k\| \quad (18)$$

$$y_k = \Omega^T * f_k, \quad y_i = \Omega^T * f_i,$$

f_k : 그림 15의 학습얼굴, f_i : 그림 13의 다



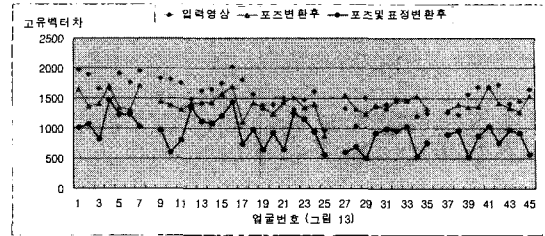
(그림 14) 포즈 및 표정 정규화 검증을 위한 얼굴 인식 실험 흐름도

양한 포즈 및 표정 영상

이때, 정확한 오차측정을 위하여 그림 16과 같이 두발 부위를 제거하고 얼굴 영역만을 추출한 영상을 사용하였다.



(그림 16) 주성분 분석을 위한 얼굴 부위 추출 예



(그림 17) 포즈 및 표정 정규화 얼굴의 고유벡터 거리차이

4.2 실험 결과 및 분석

표 2는 그림 13의 다양한 영상들을 처리한 결과 중 몇몇 예들을 중간과정과 함께 보여준 것으로서, 각 입력 얼굴 영상은 포즈 변환만을 처리한 첫 행과 포즈 및 표정을 모두 정규화 시킨 두 번째 행으로 구성되어 있다.

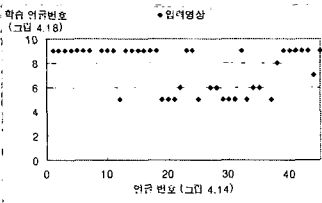
(표 2) 포즈 정규화와 포즈 및 표정 정규화

구분	입력 영상	얼굴 정합	포즈 변환 및 미러링 영역 추출	포즈 미러링
			포즈/표정 변환 및 미러링 영역 추출	포즈/표정 미러링(최종 인식영역)
그림 13의 5번 영상				
그림 13의 20번 영상				
그림 13의 25번 영상				

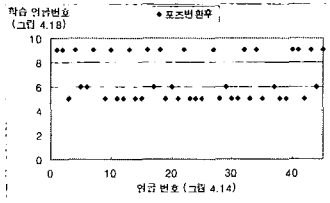
그림 17은 다양한 포즈 및 표정을 가진 그림 13의 영상을 그림 16과 같이 얼굴 부위만을 추출한 후 주성분 분석을 통하여 그림 15의 학습 얼굴 영상 중 해당 얼굴인 6번 얼굴과의 고유벡터 거리 차이값을 도식한 그래프이다. 그래프의 각 단계 거리값을 보면 어떤 영상은 최초 원 입력 영상보다 포즈만 변환한 영상이 오히려 높은 거리 차이값을 나타내었다. 이러한 영상은 대부분 눈을 크게 뜨고 있어서 눈썹이 기준치 이상으로 올라가거나 입을 크게 벌리고 있기 때문에 이를 정면 영상으로 정규화 시키는 과정에서 얼굴의 형태가 변형된 경우이다. 하지만, 표정까지 변환하여 최종 정규화된 영상은 그림 17의 동그란 표시가 보여주는 것처럼 입력 영상의 평균 고유벡터 차이값 1559.8과 포즈만 변환한 영상의 평균 고유벡터 차이값 1405.9보다 훨씬 적은 평균 922.5의 값을 보였다.

이러한 결과는 입력 영상을 그대로 인식하는 것보다는 포즈를 정규화 시킨 영상이 주성분 분석 방법과 같은 현재의 인식 시스템의 성능향상에 도움을 주며, 더 나아가서는 표정까지 분석하여 이를 교정시켜주는 것이 훨씬 효율적임을 보여준다.

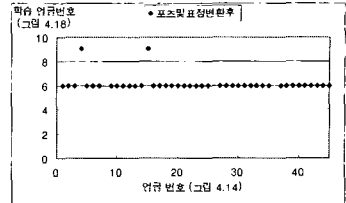
그림 18은 주성분 분석의 학습에 쓰인 그림 15의 10명과 다양한 포즈 및 표정을 가지고 있는 그림 13 영상과의 고유벡터 차이값을 비교하여 최소의 거리차이를 갖는 영상을 인식 얼굴로 판단하여 보여준 결과이다. 일반적으로 얼굴 인식을 위한 주성분 분석의 고유벡터는 유클리디언 거리를 가지고 입력 영상과 학습 영상과의 차이값을



(a) 원 영상



(b) 포즈 변환 후



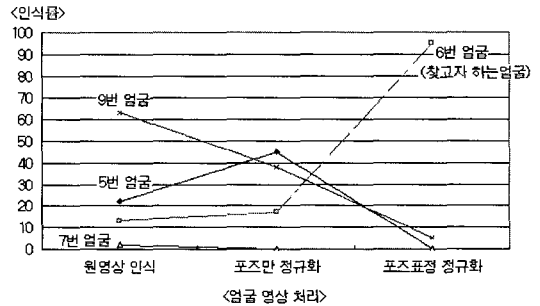
(c) 포즈 및 표정 변환 후

(그림 18) 원 영상 (그림 13)과 포즈 및 표정 변환 영상과의 인식 결과 비교

계산하며 일정한 임계값을 설정하여 이 값을 기준으로 인식 여부를 판단하지만, 본 실험에서는 정규화된 영상과 원래의 정면 영상과의 왜곡 여부를 확인하기 위한 차원이므로 오거부울을 없애기 위하여 최소값을 갖는 얼굴 영상이 유사도가 가장 높은 영상으로 판단하였다.

그림 18의 (a)는 입력 영상을 인식한 것으로서 정면에 가까운 몇 영상만이 인식하고자 하는 얼굴을 정확히 지목했을 뿐 대부분이 그림 15 (b)의 9번 영상과 5번 영상으로 분리되었다. 포즈 변환 후의 인식 결과를 살펴보면 9번 얼굴 영상의 유사도가 많이 줄어들어 찾고자 하는 6번 얼굴 영상의 인식률이 많이 늘었지만, 반면에 5번 영상에 대한 인식도 상대적으로 많이 늘었다. 이는 5번 얼굴 영상을 지목하는 분포가 널리 퍼져 있는 것으로 보아 아직까지 표정이 정규화되지 않은 영상의 얼굴 형태가 5번 얼굴 영상과 유사하다고 볼수 있다. 하지만 이러한 얼굴 영상들도 표정 정규화까지 마친 후에는 두 영상을 제외한 모든 영상이 찾고자 하는 6번 얼굴 영상과 가장 가까운 유사도를 가졌음을 그림 18 (c)의 결과를 통하여 보여주고 있다.

그림 19는 그림 18의 결과를 토대로 이를 통계적으로 분석한 결과로서, 원 영상 그대로 인식하였을 때는 과반수 이상을 차지하던 그림 15 (b)의 9번 얼굴 영상이 포즈 변환을 거쳐서는 38%로, 그리고 최종 단계인 표정 정규화를 마친 후에는 5%로 오인식률이 떨어진 것을 확인할 수 있다. 또한, 찾고자 하는 얼굴인 그림 15 (b)의 6번 얼굴 영상은 최초 원 영상에서는 13%의 인식



(그림 19) 입력 영상의 처리에 따른 얼굴 인식 분포도

률을 보이다가 포즈 정규화 과정을 거친 다음에는 17%로 상승했다. 이때 인식의 향상율이 저조한 이유는 비록 포즈를 정면 영상으로 정규화시켰지만 다양한 표정으로 인하여 찾고자 하는 얼굴 영상과의 유사도가 많이 떨어지기 때문이다. 이러한 원인은 표정 정규화 과정까지 수행한 후에야 비로소 95%의 높은 인식률을 가질 수 있었으며, 이상의 실험들을 통하여 포즈 및 표정의 정규화가 주성분 분석 방법 등을 사용하는 기존의 얼굴 인식 시스템의 인식을 향상에 기여할 수 있음을 확인할 수 있었다.

5. 결론

일반적으로 얼굴을 인식하기 위하여 분석하는 항목들 중 다양한 얼굴 포즈는 얼굴의 획득 방법에 따라 실시간으로 변형되기 때문에 얼굴을 분석하는데 있어서 커다란 문제점으로 작용한다. 따라서, 기존의 연구들은 정면 영상, 회전이 없는 정형화된 얼굴 등을 최초 입력 영상의 제약 조건으로

하여 연구를 하였으며, 현재까지 이러한 제한 사항 하에서만 높은 인식 성능을 가질 수 있었다.

이에 본 논문에서는 다양한 포즈 및 표정을 분석하여 이를 정면 영상으로 변환시키는 알고리즘을 얼굴인식 단계에 추가함으로써 기존의 인식시스템에서도 얼굴의 다양한 입력 영상을 정면 얼굴처럼 인식하고자 하는 방법을 제안하였다.

제안한 방법은 TSL 색상모델을 사용하여 입력 영상의 얼굴 영역과 얼굴의 구성요소인 양 눈과 입을 추출하고, 이를 통한 기하학적 구조를 분석하였다. 또한 분석된 구조를 이용하여 입력된 영상의 얼굴 포즈를 추정하고, 추정된 포즈를 3차원의 X축, Y축, Z축으로 성분 분해하여 이들의 조합을 통하여 정면 얼굴 영상으로 합성하였으며, 기본적인 표정의 움직임은 고려하여 표정의 정규화 방안도 제안하였다.

본 논문에서 제안한 포즈 및 표정 추정 및 정규화 얼굴로의 변환을 실험하기 위하여 다양한 각도 및 표정을 가진 얼굴 영상을 테스트하였으며, 95% 이상의 인식을 향상을 보였다. 물론, 이러한 수치는 얼굴 포즈의 각도 및 표정의 여부가 어느 정도냐에 따라 성능이 크게 좌우되겠지만 인식률의 차이값을 고려하지 않더라도 포즈 및 표정의 정규화 과정에 대한 타당성을 충분히 입증할 수 있는 실험이었다.

제안한 방법은 정확한 얼굴 특징값을 얻기 위하여 사용자에게 입력 장치의 틀에 정확히 일치하는 얼굴의 정면 영상을 요구하는 현재의 출입 관리 시스템보다는 움직이는 사람의 형체를 인식하고 이로 인하여 발생하는 얼굴의 다양한 각도 및 표정에 강인하게 사람을 판별하여 대화를 나눌 수 있는 로봇 비전 시스템 및 모바일 폰의 화상 채팅과 같은 분야에서 더 유용하게 쓰일 것으로 기대한다.

향후 연구과제로는 각도가 큰 부분의 왜곡 요인 해결 방안 및 얼굴 구성요소의 워핑에 최적화될 수 있는 전용 얼굴 모델의 설계이다.

참 고 문 헌

- [1] G. Yongsheng, "Face recognition using line edge map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764-779, 2002.
- [2] J. C. Terrillon and S. Akamatsu, "Comparative performance of different chrominance spaces for color segmentation and detection of human faces in complex scene images," *Proc. of the 4th IEEE Int'l Conf. on Automatic Face and Gesture Recognition*, pp. 54-60, 2000.
- [3] J. Ahlberg, "Model-based Coding : Extraction, Coding, and Evaluation of Face Model Parameters," *Dissertations, Dept. of Electrical Engineering, Linköping University, Sweden*, 2002.
- [4] R. Brunelli and T. Poggio, "Face Recognition: Features versus Templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042-1052, 1993.
- [5] D. J. Beymer, "Pose-Invariant Face Recognition Using Real and Virtual Views," *Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA*, 1995.
- [6] P. Ekman and W. V. Friesen, *Facial Action Coding System*, Consulting Psychologist Press, Palo Alto, CA, USA, 1997.

◎ 저자 소개 ◎



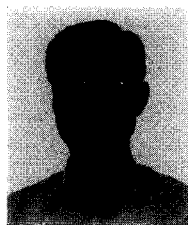
최재영

1995년 : 경원대학교 수학과 졸업(학사)
1999년 : 경원대학교 대학원 전산학과 졸업(석사)
2004년 : 경원대학교 대학원 전산학과 졸업(박사)
2004~현재 : 경원대학교 소프트웨어 연구소 책임연구원
관심분야 : 영상처리, 패턴인식, 멀티미디어, etc.
E-mail : jy1001@kyungwon.ac.kr



황보택근

1983년 : 고려대학교 공과대학 졸업(학사)
1987년 : CUNY Computer Science 졸업(석사)
1995년 : S.I.T. Computer Science 졸업(박사)
1988년~1993년 : Q-Systems, Senior Technical Staff
1995년~1997년 : 삼성종합기술원 선임연구원
1997~현재 : 경원대학교 소프트웨어대학 부교수
관심분야 : 컴퓨터 그래픽스, 영상처리, GIS, etc.
E-mail : tkwhangbo@kyungwon.ac.kr



김낙빈

1976년 : 고려대학교 전자공학과 졸업(학사)
1982년 : 고려대학교 대학원 컴퓨터공학과 졸업(석사)
1985년 : 고려대학교 대학원 컴퓨터공학과 졸업(박사)
1996년~1997년 : Univ. of Carnegie Mellon 연구교수
1986~현재 : 경원대학교 소프트웨어대학 교수
관심분야 : 멀티미디어, 디지털 영상신호 처리, 컴퓨터 구조, etc.
E-mail : nbkim@kyungwon.ac.kr