

포즈에 독립적인 얼굴 인식을 위한 얼굴 포즈 변환

정희원 박 현 선*, 박 종 일**, 김 회 율*

Face Pose Transformation for Pose Invariant Face Recognition

Hyun-Sun Park*, Jong-Il Park**, Whoi-Yul Kim* *Regular Members*

요 약

얼굴 인식 분야에서 포즈의 변화는 인식률을 저하시키는 가장 심각한 문제로 알려져 있다. 본 논문에서는 이러한 포즈가 변화된 얼굴 영상에 대한 인식률을 높이기 위한 전처리 단계로 정면이 아닌 얼굴 영상을 정면 얼굴 영상으로 변환시키는 방법을 제안한다. 제안한 방법은 PCA 계수를 선형 변환 시키는 변환 행렬을 사용되는데 이 변환 행렬은 PCA 계수 사이의 선형적인 관계를 이용하여 구한다. 제안된 방법은 PCA/LDA를 이용한 얼굴 인식 알고리즘으로 검증하였으며, 실험 결과 제안된 방법이 얼굴 인식률을 20% 정도 향상시킴을 알 수 있었다.

Key Words : face recognition, face pose, face pose transform, face pose estimation, eigenface.

ABSTRACT

Recognition of posed face is one of the most challenging problems in the field of face recognition. In this paper, as a preprocessing step for recognizing such faces, a method to transform non-frontal face images into frontal face images is proposed. The linear relationship between eigenfaces is utilized to obtain a pose transform matrix. The proposed method is verified with a well-known face recognition algorithm based on PCA/LDA. Compared to the conventional algorithm applied to the original posed face images, our experimental results indicated that the proposed method contributes to improve the recognition rate of such faces by 20%.

1. 서론

생체인식 분야에는 일반적으로 지문 인식, 홍채 인식, 얼굴 인식 등이 있다. 이 중 얼굴 인식은 지문이나 홍채 인식에 비하여 상대적으로 영상 획득이 쉽다는 장점을 지니며 무인 감시, 비디오 분석, 엔터테인먼트와 같은 다양한 분야에서 응용되고 있다. 그러나 이러한 다양한 응용 분야를 갖는 얼굴 인식은 그 성능이 포즈나 조명 변화에 매우 민감하다는 단점을 가지고 있다. 이 중 특히 포즈 변화를 일으키는 얼굴의 회전은 얼굴 인식률을 저하시키는 데에 매우 심각한 요인으로 작용하고 있다¹⁾.

얼굴 영상의 회전은 3차원 상의 어느 축을 중심

으로 회전하는가에 따라 그림 1과 같이 세 가지의 경우로 나뉘어 질 수 있다. z축을 중심으로 회전한 영상의 경우 단순한 2차원 회전 변환을 이용하여 정상적인 영상으로 변환이 가능하다. 여기에서 회전 각도를 알아내야 하는 문제점이 있기는 하지만 눈을 검출하는 다양한 알고리즘²⁻⁴⁾이 개발되어 있기 때문에 눈을 검출한 다음 눈의 좌표를 이용하여 회전시킬 수 있다.

얼굴인식에서 가장 문제가 되는 회전은 x축과y축에 대한 회전이다. 이런 회전은 affine transform을 이용하더라도 정면 얼굴로의 변환이 불가능하다. x축과 y축에 대한 회전을 해결하기 위해서 얼굴의 3차원 모델링을 통하여 얼굴을 인식하는 방법⁵⁻⁷⁾이

* 한양대학교 전자통신전파공학과 영상공학연구소(wykim@hanyang.ac.kr), ** 한양대학교 전자통신전파공학과 가상현실연구소
논문번호 : KICS2004-07-122, 접수일자 : 2004년 7월 29일

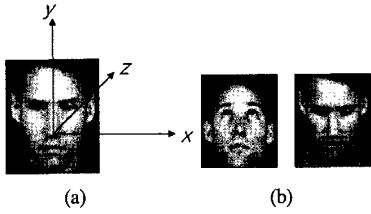


그림 1. 얼굴 회전 (a) 얼굴 회전축, (b) x축 중심 회전, (c) y축 중심 회전, (d) z축 중심 회전

제안되었지만 얼굴의 특징점인 눈, 코, 입 등의 위치와 얼굴 윤곽 등 매우 다양한 특징점을 정확히 파악하여야 하고 계산이 복잡한 문제점이 있다.

본 논문에서는 x축과 y축에 대하여 회전된 얼굴을 정면 얼굴로 변화하기 위하여 통계적인 방법을 이용하여 얼굴 변환 행렬을 구한 후 이를 이용하여 정면 얼굴로 변환하는 방법을 제안한다. 본 논문에서 이용한 통계적인 방법이란 미리 정해진 트레이닝 셋에서 포즈 변화가 있는 얼굴과 정면 얼굴의 PCA 계수의 상관관계를 형성하는 선형 방정식의 유도를 통하여 포즈 변환 행렬을 구한 다음, 이 포즈 변환 행렬을 이용하여 포즈 변화가 있는 얼굴 영상을 정면 얼굴 영상으로 변환하는 것이다. 이러한 통계적인 방법은 트레이닝 시에 대용량의 데이터를 처리해야 하기 때문에 계산이 복잡하고 메모리 요구가 매우 높지만, 실제 얼굴 포즈 변환 단계에서는 단순한 행렬 곱만이 필요하기 때문에 계산 복잡도의 증가는 미미하다.

본 논문의 구성은 제 2 장에서 포즈 변환을 위한 선결 과제인 포즈 분류에 대한 알고리즘을 설명하고, 제 3 장에서는 본 논문에서 제안된 포즈 변환 알고리즘을 설명한다. 제 4 장에서는 실험에 이용된 데이터 셋에 대한 설명과 실험 결과를 보여주고, 제 5 장에서 결론을 맺는다.

II. 포즈 분류

포즈 변환을 하기 위해서 제일 먼저 해야 할 일이 포즈 분류다. 포즈 분류는 입력되는 얼굴의 회전 방향을 분류하는 것이다. 본 논문에서 이용한 포즈 분류 방법은 입력 얼굴 영상의 PCA 계수의 mahalanobis distance를 이용하였다. 이는 포즈 변환 단계에서 이용하는 PCA를 재사용하기 위해서이다.

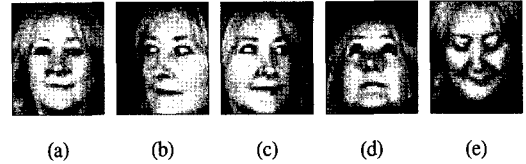


그림 2. 실험에 사용된 얼굴 영상의 포즈들 (a) F 포즈, (b) R 포즈, (c) L 포즈, (d) U 포즈, (e) D 포즈

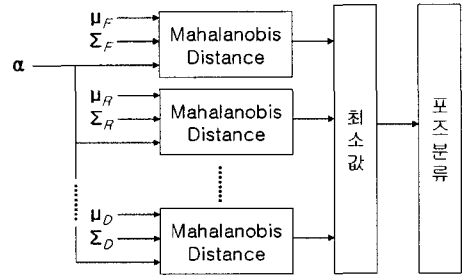


그림 3. 포즈 분류 알고리즘

2.1 포즈 분류 알고리즘

본 논문에서 사용한 트레이닝 셋의 포즈는 그림 2와 같이 일정한 각도로 회전된 다섯 가지의 포즈로 분류되었다.

포즈를 분류하는 알고리즘은 그림 3과 같이 입력된 고유 얼굴에 투영된 계수 값을 입력으로 받아 각 포즈에 해당하는 평균 벡터와 공분산 행렬을 이용하여 식 (1)과 같이 mahalanobis distance를 계산한다. 계산된 거리 중에서 가장 최소가 되는 포즈를 분류된 포즈라고 본다.

$$EstimatedPose = \arg \min_{p=F,R,L,U,D} \sqrt{(\alpha - \mu_p)' \Sigma_p^{-1} (\alpha - \mu_p)} \quad (1)$$

여기서, α 는 PCA 계수이고, μ_p 는 p(=R, L, U, D) 포즈 얼굴의 평균 얼굴, Σ_p 는 p 포즈 얼굴 벡터의 공분산 행렬이다.

포즈 분류 알고리즘을 이용하여 테스트 셋에 적용한 결과 포즈 측정의 정확도는 표 1과 같다. 실험에 사용된 테스트 셋은 각 포즈 별로 296 개의 샘플을 가지고 있다.

표 1. 포즈 분류의 정확도

포즈	총 샘플 개수	오류 샘플 개수	정확도 (%)
F	296	21	92.9
R	296	5	98.3
L	296	11	96.3
U	296	22	92.3
D	296	3	99.3
평균	1480	62	95.8

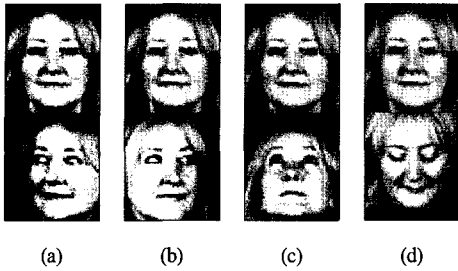


그림 4. 포즈 쌍 (a) FR 포즈 쌍, (b) FL 포즈 쌍, (c) FU 포즈 쌍, (d) FD 포즈 쌍

III. 포즈 변환

본 논문에서는 회전된 얼굴을 정면 얼굴로 변화하기 위하여 통계적인 방법을 이용하여 얼굴 변환 행렬을 구한 후 이를 이용하여 정면 얼굴로 변환하는 방법을 제안하였다. 본 논문에서 이용한 통계적인 방법이란 미리 정해진 트레이닝 셋에서 포즈 변화가 있는 얼굴과 정면 얼굴의 PCA 계수의 상관관계를 형성하는 선형 방정식의 유도를 통하여 포즈 변환 행렬을 구한 다음, 이 포즈 변환 행렬을 이용하여 포즈 변화가 있는 얼굴 영상을 정면 얼굴 영상으로 변환하는 것이다. 이를 위한 포즈 변환 행렬은 식 (2), (3), (4), (5)에 나와 있다.

본 논문에서는 계산의 편의를 위해서 F 포즈와 R, L, U, D 포즈의 얼굴을 그림 4와 같이 쌍으로 만들었다.

본 논문에서 사용한 포즈 변환 행렬을 위한 트레이닝 셋은 (F, R, L, U, D, F, R, L, U, D)/1명 × 147명 = 1470 개의 영상으로 되어 있다. 따라서, 각 포즈 쌍 별로 생성될 수 있는 트레이닝 셋의 개수는 (2×2 포즈 쌍)/1명×147명 = 588 개가 된다. 또한 포즈 변환에 사용되는 Principal component의 개수는 90 개를 사용하였다.

각 포즈의 변환 행렬을 A_p ($p = R, L, U, D$)라고 하면 행렬 A_p 는 식 (2), (3), (4), (5)와 같은 관계를 나타낸다.

- RF 변환 행렬 A_R

$$\begin{aligned} C_F &= A_R C_R \\ A_R &= C_F C_R^+ \end{aligned} \tag{2}$$

--LF 변환 행렬 A_L

$$\begin{aligned} C_F &= A_L C_L \\ A_L &= C_F C_L^+ \end{aligned} \tag{3}$$

- UF 변환 행렬 A_U

$$\begin{aligned} C_F &= A_U C_U \\ A_U &= C_F C_U^+ \end{aligned} \tag{4}$$

- DF 변환 행렬 A_D

$$\begin{aligned} C_F &= A_D C_D \\ A_D &= C_F C_D^+ \end{aligned} \tag{5}$$

여기서 C_F 와 C_p ($p = R, L, U, D$)는 식 (6)과 그림 5와 같이 각각 F, R, L, U, D 포즈에 해당하는 모든 트레이닝 셋의 PCA 계수를 행벡터로 하는 행렬이며, C^+ 은 행렬 C 의 pseudo inverse matrix이다. 여기서 pseudo inverse matrix를 이용한 이유는 행렬 C 의 차원이 588×90 으로 정방 행렬이 되지 않아서 일반적인 역행렬을 구할 수 없기 때문이다.

$$\begin{bmatrix} X_{F,1}^1 & X_{F,2}^1 & \dots & X_{F,K}^1 \\ X_{F,1}^2 & X_{F,2}^2 & \dots & X_{F,K}^2 \\ \vdots & \vdots & \vdots & \vdots \\ X_{F,1}^M & X_{F,2}^M & \dots & X_{F,K}^M \end{bmatrix} = A_p \begin{bmatrix} X_{p,1}^1 & X_{p,2}^1 & \dots & X_{p,K}^1 \\ X_{p,1}^2 & X_{p,2}^2 & \dots & X_{p,K}^2 \\ \vdots & \vdots & \vdots & \vdots \\ X_{p,1}^M & X_{p,2}^M & \dots & X_{p,K}^M \end{bmatrix} \tag{6}$$

단, 여기서 $p=(R, L, U, D)$, M 은 트레이닝 샘플의 개수, K 는 고유 얼굴의 개수이다.

얼굴 포즈를 변환하는 전체 알고리즘은 그림 6과 같다. 입력된 얼굴 영상은 고유 얼굴 공간에 투영되고 이 투영된 계수 값 중에서 가장 지배적인 20개의 계수 값을 가지고 식 (1)을 이용하여 포즈를 분류한다. 분류된 포즈에 따라 각 포즈에 해당하는 포

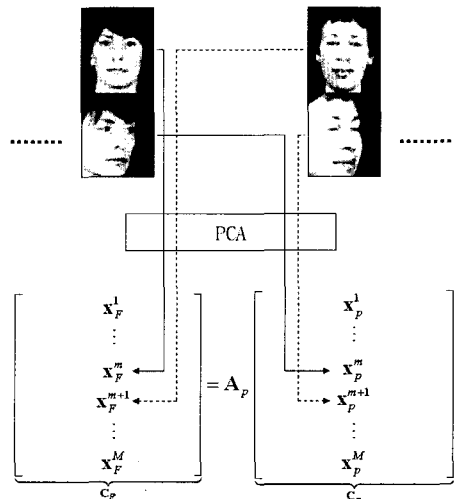


그림 5. 포즈 쌍을 이용하여 C_F 와 C_p 행렬 만들기

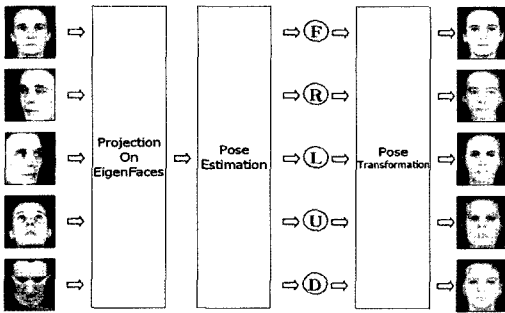


그림 6. 전체 시스템 개요도

즈 변환 행렬을 이용하여 계수 값을 선형 변환한 후 복원한다.

IV. 실험 결과

4.1 실험 데이터 셋

실험에 쓰인 데이터 셋은 MPEG-7에서 얼굴 인식 기술자의 성능 평가를 위한 데이터 셋 중에서 포즈의 변화가 가장 심한 xm2vt 데이터 셋으로 총 (F, R, L, U, D, F, R, L, U, D)/1명X295명=2950 개의 샘플로 이루어져 있다. 영상의 크기는 46x56 이며 눈의 위치가 (16,24), (31,24)로 고정되어 있다.

전체 데이터 셋 중에서 147 명은 포즈 변환 행렬 생성을 위한 트레이닝 셋으로, 나머지 148 명은 생성된 포즈 변환 행렬의 타당성 검증을 위한 테스트 셋으로 분류 하였다. 그림 7은 트레이닝 셋을 그림 8은 테스트 셋의 일부를 보여주고 있다.



그림 7. 실험에 쓰인 트레이닝 셋



그림 8. 실험에 쓰인 테스트 셋

4.2 포즈 변환 결과

제안된 포즈 변환 방법을 이용하여 트레이닝 셋과 테스트 셋에 적용하여 보았다. 결과는 그림 9, 그림 10, 그림 11과 같다.

그림 9는 트레이닝 샘플이 어떻게 변환되는가를 보고 포즈 변환 행렬이 정상적으로 만들어졌는가를 알아보기 위한 그림이다. 그림에서 보는 바와 같이

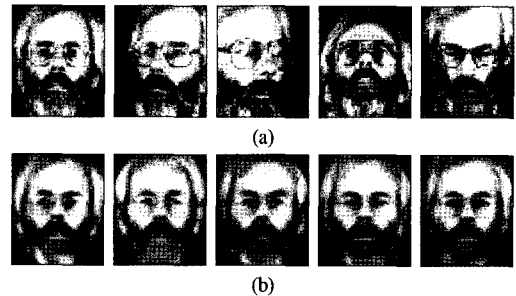


그림 9. 트레이닝 셋의 포즈 변환 (a) 원영상, (b) 포즈 변환 영상

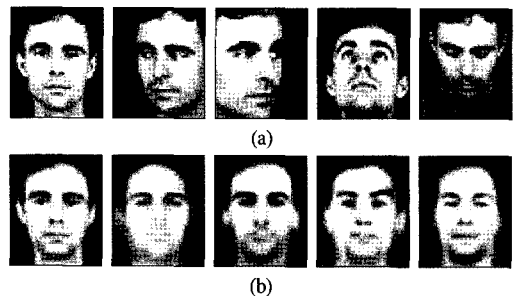


그림 10. 테스트 셋의 포즈 변환(1) (a) 원영상, (b) 포즈 변환 영상

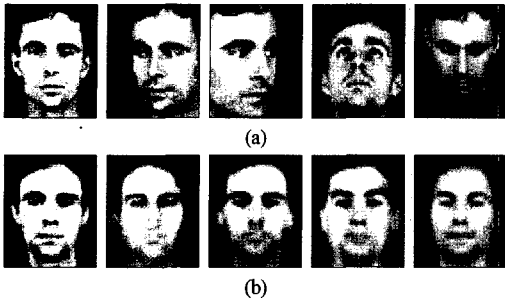


그림 11. 테스트 셋의 포즈 변환(2) (a) 원영상, (b) 포즈 변환 영상

매우 잘 만들어져 있는 것을 알 수 있다. 그림 10, 그림 11은 테스트 샘플에 포즈 변환 행렬을 적용한 예이며, 그 결과를 보면 매우 근사하게 변환되는 것을 알 수 있다. 포즈 변환 행렬의 타당성 검증을 위하여 본 논문에서는 PCA/LDA^[1, 9-10]를 이용하여 얼굴 인식을 해보았다.

4.3 PCA/LDA를 이용한 포즈 변환의 검증

포즈 변환 행렬을 이용한 포즈 변환의 검증을 위하여 PCA/LDA를 이용하여 얼굴 인식을 하였다. 이 때 PCA에 쓰인 기저의 개수는 90 개이며, LDA에 쓰인 기저의 개수는 48 개로 하였다. 실험 과정은 그림 12에 나와 있는 것처럼 포즈 변환을 하지 않고 얼굴 인식을 수행하는 실험 1과 포즈 변환을 통하여 정면 얼굴로 변환된 이미지를 가지고 얼굴 인식을 수행하는 실험 2로 구분하였다.

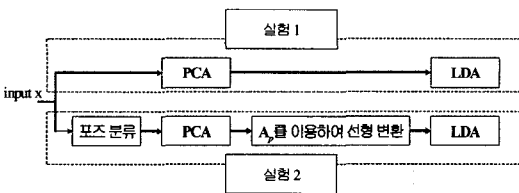


그림 12. 포즈 변환의 검증을 위한 얼굴 인식 실험

얼굴 인식률을 알아보기 위하여 본 논문에서는 MPEG-7에서 정의된 ANMRR (Average Normalized Modified Retrieval Rank) 척도를 사용하였다. ANMRR에 대하여 설명하기 전에 다음 기호의 정의에 대해서 알아보자^[11, 12].

- $NG(q)$: 쿼리 q 에 대한 ground truth image 개수
- $K(q) = \min(\min(4 \times NG(q), 2 \times GTM),$
- 여기서 $GTM = \max\{NG(q)\}$ for all q 's.

- $R(k)$: 검색 결과에서 이미지 k 의 순위

$Rank(k)$ 는 식 (7)과 같이 정의한다.

$$Rank(k) = \begin{cases} R(k) & \text{if } R(k) \leq K(q) \\ (k+1) & \text{if } R(k) > K(q) \end{cases} \quad (7)$$

식 (7)을 이용하여 쿼리 q 에 대한 AVR (Average Rank)는 식 (8)과 같이 정의한다.

$$AVR(q) = \sum_{k=1}^{NG(q)} \frac{Rank(k)}{NG(q)} \quad (8)$$

여기서 $AVR(q)$ 가 $NG(q)$ 에 종속되어 있기 때문에 이 영향을 줄여주기 위하여 MRR (Modified Retrieval Rate)을 식 (9)와 같이 정의한다.

$$MRR(q) = AVR(q) - 0.5 - \frac{NG(q)}{2} \quad (9)$$

MRR을 정규화한 NMRR (Normalized Modified Retrieval Rate)은 식 (10)과 같이 정의한다.

$$NMRR(q) = \frac{MRR(q)}{K + 0.5 - 0.5 \times NG(q)} \quad (10)$$

NMRR은 0부터 1사이의 값을 갖는데 0에 가까울수록 검색 효율이 좋은 것이다. 모든 쿼리 셋 Q 에 대하여 평가된 NMRR의 평균 ANMRR은 식 (11)과 같다.

$$ANMRR(q) = \frac{1}{Q} \sum_{q=1}^Q NMRR(q) \quad (11)$$

표 2는 얼굴 인식 실험 결과를 보여주고 있다. 이 때, 얼굴 인식 실험을 위한 데이터 셋의 구분은 그림 13에 나와 있는 것처럼 포즈 변환 행렬 생성에 쓰인 샘플들을 제외한 148 명의 샘플 중에서 74 명을 트레이닝 셋으로, 그리고 나머지 74 명을 테스트 셋으로 하였다.

표 2. 얼굴 인식 실험 결과

	ANMRR
실험 1	0.549807
실험 2	0.421403

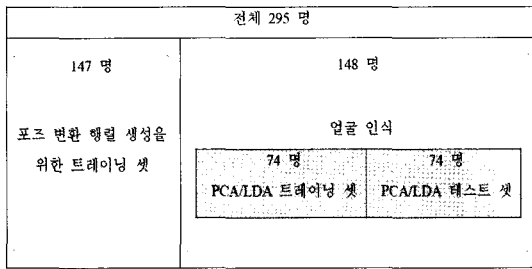


그림 13. 실험에 쓰인 데이터 셋의 분류

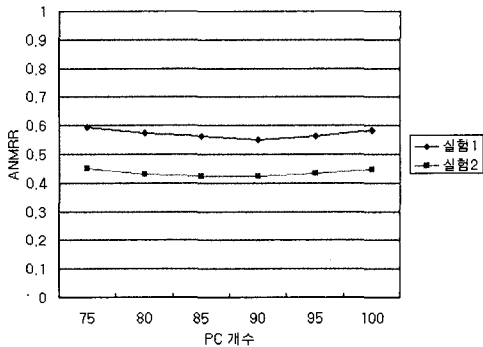


그림 14. Principal Component 개수에 따른 인식률 비교

실험 결과를 보면 원 영상에 대한 얼굴 인식률(실험 1)보다 포즈를 정면으로 변환한 후 수행한 얼굴 인식률(실험 2)이 훨씬 더 높음을 알 수 있다.

그림 14는 포즈 변환 행렬을 만들 시 사용되는 Principal Component의 개수에 따른 최종 얼굴 인식률을 보여주고 있다. 그림에 보듯이 Principal Component의 개수가 90 개일 때 가장 우수한 성능을 보여주고 있음을 알 수 있다.

IV. 결론 및 향후 과제

본 논문에서는 얼굴 인식에서 가장 큰 문제점으로 알려져 있는 얼굴 포즈의 변화에 대한 인식률 저하 문제를 해결하기 위한 포즈 변환 방법을 제안하였다. 포즈 변화가 있는 얼굴과 정면 얼굴의 PCA 계수를 이용한 선형적인 상관 관계를 분석하고 이를 통한 포즈 변환 행렬을 유도하였다. 유도된 포즈 변환 행렬을 이용하여 정면 얼굴을 생성함으로써 인식률 저하 요인을 발생시키는 얼굴 회전에 의한 특징 값의 변화 문제를 해결하였다.

얼굴 인식률 측정을 위하여 PCA/LDA 알고리즘을 이용한 인식률 결과를 통하여 제안한 방법의 성능 향상 효과를 입증하였다. 또한, 본 논문에서 제

안한 알고리즘은 기존의 3차원 모델링을 통한 얼굴 회전을 이용한 방식에서 문제가 되는 특징점 추출 및 3차원 얼굴 모델링의 높은 계산 복잡도와 이로 인한 구현의 문제에 대한 어려움을 극복하여 실제 시스템에 간단히 적용할 수 있을 것으로 판단된다.

앞으로 좀더 향상된 인식률을 얻기 위한 포즈 변환 행렬의 개선을 위해서 비선형성을 고려한 포즈 변환 행렬 생성 방법 및 얼굴 요소 별 포즈 변환 행렬에 대하여 연구해야 할 것이다.

참고 문헌

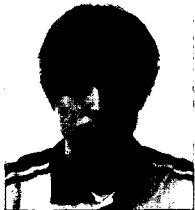
- [1] W. Zhao, R. Chellappa, P. J. Phillips, A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys (CSUR)*, Vol. 35, No. 4, pp. 399-458, 2003.
- [2] 지형근, 이경희, 정용화, "에지 정보와 SVM의 결합을 통한 눈 검출", *대한전자공학회 하계 종합학술대회*, pp. 347-350, 2002.
- [3] H. Tani, K. Terada, S. Oe, J. Yamaguchi, "Detecting of one's eye from facial image by using genetic algorithm," *IECON*, Vol. 3, pp. 1937-1940, 2001.
- [4] R. T. Kumar, S. K. Raja, A. G. Ramakrishnan, "Eye detection using color cues and projection functions," *Int. Conf. on Image Processing*, Vol. 3, pp. 337-340, 2002.
- [5] Chongzhen Zhang, Fernald S. Cohen, "3-D Face Structure Extracting and Recognition From Images Using 3-D Morphing and Distance Mapping," *IEEE Trans. on Image Processing*, Vol. 11, No. 11, pp. 1249-1259, November 2002.
- [6] V. Blanz, S. Romdhani, T. Vetter, "Face Identification across Different Poses and Illuminations with a 3D Morphable Model", *Int. Conf. on Automatic Face and Gesture Recognition*, pp. 202-207, 2002.
- [7] A-Nasser Ansari, Mohamed Abdel-Mottaleb, "3-D Face Modeling Using Two Views and a Generic Face Model with Application to 3-D Face Recognition," *IEEE Conf. on Advanced and Video and Signal Based Surveillance*, pp. 37-44, 2003.
- [8] M.A. Turk, A.P. Pentland, "Face Recognition

Using Eigenfaces," *IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

- [9] K. Etemad, R. Chellappa, "Discriminant Analysis for Recognition of Human Face Images," *Journal of the Optical Society of America*, Vol. 14, pp. 1724-1733, 1997.
- [10] W. Zhao, R. Chellappa, A. Krishnaswamy, "Discriminant Analysis of Principal Components for Face Recognition," *Int. Conf. on Automatic Face and Gesture Recognition*, pp. 336-341, 1998.
- [11] B. S. Manjunath, Jens Rainer Ohm, Vinod V. Vasudevan, Akio Yamada, "Color and texture descriptors," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 11, No. 6, pp. 703-715, June 2001.
- [12] V. V. Vinod and B. S. Manjunath, "Report on AHG of Color and Texture," *ISO/IEC/JTC1/SC29/WG11, Doc. M5560*, December 1999.

박 현 선 (Hyun-Sun Park)

정회원



2001년 2월 한양대학교 전자
전기공학부(공학사)
2003년 2월 한양대학교 대학
원 전자통신전파공학과(공학
석사)
2004년~현재 LG전자 연구원
<관심분야> 패턴인식, 얼굴인식

등.

박 종 일 (Jong-Il Park)

정회원



1987년 서울대학교 전자공학과
(공학사)
1989년 서울대학교 대학원 전자
공학과(공학석사)
1995년 서울대학교 대학원 전
자공학과(공학박사)
1992년~1994년 일본 NHK 방송
기술연구소 객원연구원
1995년~1996년 한국방송개발원 선임연구원
1996년~1999년 일본 ATR 지능영상통신연구소 연구원
1999년~현재 한양대학교 전자전기컴퓨터공학부 교수
<관심분야> 가상현실, 컴퓨터그래픽스/비전, 3차원 영
상처리, 인간컴퓨터 상호작용 등

김 희 울 (Whoi-Yul Kim)

정회원



1980년 한양대학교 전자공학과
(공학사)
1983년 Pennsylvania State Uni-
versity 전기공학과 (공학석사)
1989년 Purdue University 전기
공학과(공학박사)
1989년 9월~1994년 2월 Univer-
sity of Texas 조교수

1994년~현재 한양대학교 전자전기컴퓨터공학부 정
교수
<관심분야> 영상처리, 컴퓨터비전, 패턴인식, 머신비
전, MPEG-7 등.